Mathematical modelling of fibroblasts in cancer

Esther Wershof

Imperial College London

and

The Francis Crick Institute

A thesis submitted for the degree of

Doctor of Philosophy

Imperial College London

September 2019

Declaration

I Esther Wershof confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Copyright Declaration

Abstract

Cancer-associated fibroblasts (CAFs) and the associated extracellular matrix (ECM) constitute a significant part of the tumour microenvironment (TME), playing an important role in the invasive potential of the tumour. The alignment of CAFs and the corresponding ECM which they produce and organise is linked with increased cancer invasion. Additionally, massive variation in the physical architecture of the ECM is observed in both normal and pathological tissues for example swirling, diffuse or porous patterns. How these mesoscale patterns arise remains largely unexplored.

An agent-based flocking model was developed to investigate CAF properties and their involvement in emergent alignment. The model established that aligning cells had a requirement of highly persistent migration coupled with an active cell-cell collision guidance mechanism. The model predicted that alignment was a fragile state which could be easily destroyed in a heterogeneous population. These findings were confirmed experimentally.

The model was then extended to include a second underlying layer of ECM fibres that the CAFs could produce, degrade and rearrange but were also instructed to follow, constituting a CAF-ECM feedback loop. This mechanism was capable of generating diverse matrix patterns, reminiscent of those seen *in vivo*. The model was challenged to unpick the process of interconversion between matrix patterns as seen in cancer, wound healing and ageing, which it elucidated with considerable success.

Finally, clinical samples of ECM were quantified to establish if certain metrics of ECM architecture could be useful clinical prognostic factors. Early results suggest this to be true. Matrix patterns were quantified by a carefully constructed software pipeline suitable for use by other researchers on versatile data samples.

This thesis examines the causes and consequences of different ECM architectures. It is hoped that these findings will pave the way for further research into manipulating tumours towards patterns favourable for inhibiting cancer invasion.

Acknowledgements

Firstly, I would like to express my profound gratitude to all of the individual donors to Cancer Research UK and the many patients who agree to participate in research. Their courage and generosity are causing very real and exciting advances in cancer treatment. Amazing progress has been made and will continue to be made thanks to them.

My thanks go to Ruth Baker and Robert Endres for reading my work and allowing me to defend this thesis.

I am immensely grateful to Paul Bates for giving me the chance to know the wonderful world of research. He has given me the space to learn and explore my own ideas and his relentless enthusiasm and wonder for science have been infectious and uplifting. It has been a huge pleasure to work with him and follow his example of fearlessness and excitement in science. I hope that my future supervisors and colleagues will be as kind and generous as he is.

It has been a privilege to learn from Erik Sahai, whose guidance has greatly shaped my research and development as a scientist. He is a truly brilliant biologist and a wonderful role model. His ability to think creatively and draw links between scientific ideas are remarkable and his dedication to science is inspiring.

I would like to say a huge thank you to my collaborator Danielle Park for her patience. She showed me the ropes when I arrived knowing nothing at all and has taught me the value of collaboration. I would not have been able to make anywhere near as much progress without her.

My gratitude also goes to Xiao Fu for his guidance, Raphaël Chaleil for keeping everything working, and all the other members of the Bates group for making such an enjoyable working environment. Thank you also to all members of the Sahai group, in particular to Robert Jenkins for all his help and wisdom.

It has been an honour to work at The Francis Crick Institute. I have been given opportunities I could never have imagined. Coming in every day to work with such brilliant scientists in such a beautiful building fills me with immense pride. I feel incredibly privileged to be part of the Crick, an institute which will no doubt go from strength to strength with the hard work and vision of Paul Nurse, Sam Barrell and all those who work to make the Crick the amazing place it is.

Finally, I would like to thank my family: Lionel, Mum and Dad. They have been there with me every step of the way, encouraging me and looking after me. They are my cheerleaders and have faith in me even when I lose it in myself. Their love and support fill my life with joy and pick me up when I need it most. This thesis is dedicated to them.

Table of Contents

Table of figures

List of tables

Abbreviations

CAFs	Cancer-associated fibroblasts
TME	Tumour microenvironment
ECM	Extracellular matrix
αSMA	α-smooth muscle actin
MMPs	Matrix metalloproteinases
EMT	Epithelial-mesenchymal transition
TACS	Tumour-associated collagen signature
CIL	Contact inhibition of locomotion
CNN	Convolutional neural network
PRW	Persistent random walk
SEM	Subcellular element model
CPM	Cellular Potts model
LRA	Long-range alignment
SRA	Short-range alignment
HDM	High-density matrix
Curv	Curvature
Frac	Fractal dimension
API	Application programming interface
PDGF	Platelet-derived growth factor
PDGFR	Platelet-derived growth factor receptor
FUD	Functional upstream domain
PCA	Principal component analysis
t-SNE	t-stochastic neighbour embedding
depRate	Deposition rate
degRate	Degradation rate
reRate	Rearrangement rate
H&E	Hematoxylin and eosin
PSR	Picosirius red
PCK	Pan cytokeratin
TS	Tumour size
TG	Tumour grade

NBP	Normalised number of branch points
NEP	Normalised number of end points
Lac	Lacunarity

Chapter 1. Introduction

Examples of collective behaviour are ubiquitous in biology (Battersby, 2015). From the response of a shoal of thousands of fish to a predator, to the beautiful murmuration of starlings, the collective response is so much greater than the sum of its parts. In our bodies made up more than one trillion cells, a single cell has an almost negligible effect. And yet the remarkable synergy of collective cell behaviour can heal our wounds or cause us to develop perfectly formed fingers and toes. Collective cell responses also underpin unwanted phenomena, such as cancer invasion. It is therefore of crucial importance to understand how cells coordinate their behaviour. Mathematical modelling is an indispensable tool for understanding how interactions between cell neighbours give rise to emergent behaviour over huge distances and millions of cells (Lukeman, Li and Edelstein-Keshet, 2010).

1.1 Cancer evolution

The daunting challenge of curing and preventing cancer lies in the enormous diversity and adaptability of the disease. This diversity occurs on three levels: from patient-to-patient, the spatial location in the body and through intra-tumoural heterogeneity. Cancer cells evolve with the potential to acquire an infinite combination of somatic mutations. Much like Darwin's theory of evolution and the range of animals seen today, this process leads to "survival of the fittest" and a huge spectrum of surviving families of cells, each with unique characteristics and mutations.

Much work is being done to identify and treat frequent driver mutations such as BRAF mutations (Jamal-Hanjani *et al.*, 2017). This is an exciting area of research in which many important discoveries are being made due to enormous advances in computational power (McGranahan and Swanton, 2017). These improvements in understanding of tumour evolution are likely to be of deep importance in the improvement of cancer treatment.

However, the staggeringly high degree of heterogeneity in the tumour presents a significant problem to the precision medicine approach that has been promoted in recent years (McGranahan and Swanton, 2015). Within a tumour there are millions of heterogeneous cells (Quaranta *et al.*, 2009). Multiple tumour biopsies are only able to offer a snapshot of certain parts of a tumour in a state of constant, rapid spatial and temporal change. It is therefore a formidable task to try and treat all potential subpopulations. Given this immense tumour heterogeneity, are there any unifying elements common to all cancers, to circumvent these problems?

Every single tumour exists as a unique society consisting not only of cancer cells but of many other cell and structure types (Jiao and Torquato, 2011; Augsten, 2014; Heindl, Nawaz and Yuan, 2015). These other parts of the tumour micro-environment, which are not the cancer cells are known as the tumour stroma (Figure 1). These stromal cells, whilst under the influence of cancer cells, are not themselves cancerous and so are not evolving and acquiring somatic mutations like the cancer cells. Typically, cancer cannot progress without a stroma, a supporting environment with which it can develop and co-evolve. Some cancers consist of more stroma than cancer cells. In fact, stromal cells (in particular, fibroblasts) may make up around 90% of tumour cells in pancreatic cancers (Karagiannis *et al.*, 2012) rendering the tumour stroma a prime target for cancer research.

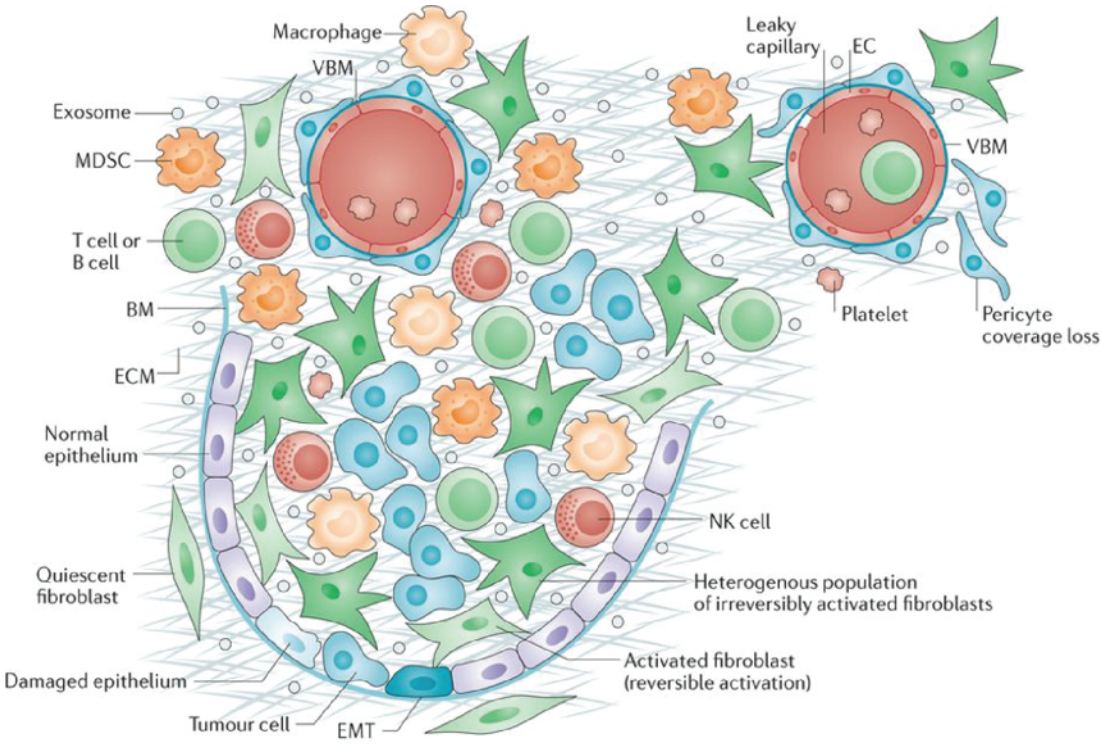


Figure 1: A schematic of the tumour microenvironment.

Schematic of the tumour microenvironment containing many cell types and other material. Fibroblasts may be quiescent or activated in either a reversible or irreversible state. Signalling in wound healing causes fibroblasts to become activated. Upon cessation of the signalling, the fibroblasts revert to their inactive state (reversible). In pathologies such as cancer fibroblasts can become irreversibly activated. Irreversible fibroblast activation is associated with enhanced αSMA expression and are then known as cancer-associated fibroblasts (CAFs). Fibroblasts constitute a large volume of the tumour and are heterogeneous. EMT=Epithelial-to-mesenchymal transition, VBM=vascular basement membrane. Image reproduced with permission from (Kalluri, 2016).

Currently, research on the tumour microenvironment can be broadly classified into two areas. The first is immunotherapy, using the body's immune system to orchestrate an attack on the cancer cells (Gotwals *et al.*, 2017). This profound idea of harnessing the power of the immune cells to identify and attack cancer cells can potentially circumvent the problem of cancer cell heterogeneity. Immunotherapy has already produced some astounding results (Feig *et al.*, 2013; Joyce and Fearon, 2015; Jin *et al.*, 2016; Gotwals *et al.*, 2017) and is likely to continue to improve cancer treatment in the future.

The second is through manipulating the stroma directly. By understanding spatial and temporal development of the stroma, explaining collective behaviour (Trepat *et al.*, 2009; Mayor and Etienne-Manneville, 2016) and identifying how the stroma is helping the cancer, the tumour and metastasis can be inhibited (Shimoda, Mellody and Orimo, 2010). It is this area which constitutes the focus of this work, with an emphasis on cancer-associated fibroblasts (CAFs), one of the most abundant stromal cell-types (Augsten, 2014). CAFs can constitute a large part of the stroma and have a complex relationship with the cancer cells (Kalluri and Zeisberg, 2006; Augsten, 2014; Kalluri, 2016). In some instances, CAFs can aid metastasis and in others, inhibit it (Özdemir *et al.*, 2014). Certain CAF populations can produce remarkable collective patterns *in vivo* such as alignment. These high-order patterns have potential implications for metastasis (Provenzano *et al.*, 2006; Riching *et al.*, 2015). This leads to the first of three principle research questions for this thesis:

Research Question 1: What properties of CAFs cause different collective behaviours?

1.2 Evolution of a tumour from start to finish

The human body contains trillions of cells and every day billions of cells are made and die. Our DNA could stretch to the sun and back around 250 times. Given the spectacular complexity of cells, it is remarkable that so many cells manage to divide without acquiring significant genetic faults. This is largely due to molecular repair mechanisms, which identify and fix molecular damage that has occurred in the cells (Lindahl, 1974). Other faulty cells induce apoptosis. Mutated cells which do not die as quickly as they divide form the origins of cancer. These cells can acquire more mutations, constantly dividing and forming the basis of a tumour. The excessive division of cancer cells results in competition for resources, leading to hypoxia (lack of oxygen) resulting from inadequate blood supply (Bearer *et al.*, 2009; Madsen *et al.*, 2015; Carmona-Fontaine *et al.*, 2017). This lack of resources can cause necrosis in the tumour centre (Gatenby and Gawlinski, 2003; Valastyan and Weinberg, 2011). Given the heterogeneity of mutated cancer cells, some subpopulations may acquire the ability to survive in these harsher environments.

Further, treatment with chemotherapy and radiotherapy can also cause further selective pressures on cancer cells (Hirata and Sahai, 2017).

Simultaneously, the cancer cells begin recruiting stromal cells to establish the tumour microenvironment (TME). Through chemical signalling such as PDGF, VEGF and TGFβ, other cells such as fibroblasts, T-cells and blood cells are recruited to the tumour (Lynch *et al.*, 1987; Karagiannis *et al.*, 2012; Augsten, 2014). These stromal cells interact and evolve with the cancer cells, inextricably linking the cancer cells with the TME and inducing a complex network of feedback loops in the tumour. Fibroblasts recruited to the tumour are largely responsible for the production and maintenance of extracellular matrix (ECM) (Kalluri and Zeisberg, 2006; Kalluri, 2016).

The ECM is a network of fibres composed primarily of collagen and fibronectin, which helps give structure to the body and is essential in wound-healing. The ECM functions both chemically and physically (Pizzo *et al.*, 2005; Vedula *et al.*, 2012). At the chemical level, the ECM released cytokines that alter cell behaviours. At the physical level, the ECM fibres act as a physical barrier inducing changes in the cells. One such example is in the feed forward loop of YAP transcription factor activation (Calvo *et al.*, 2013). YAP is required for CAFs to promote ECM stiffening. At the same time, matrix stiffening enhances YAP activation in cells. Another example is the physical constraint of moving through the ECM causing cells to change their shape, phenotype and mode of migration between amoeboid and actin-driven protrusions. (Tozluoğlu *et al.*, 2013a). Cells are able to use the ECM fibres to migrate along and cells often migrate up a stiffness gradient (durotaxis) or a chemical concentration gradient (chemotaxis) (Schlüter, Ramis-Conde and Chaplain, 2012; Ahmadzadeh *et al.*, 2017; van Helvert, Storm and Friedl, 2018). This conversion of physical signals to an intrinsic chemical change in cells is known as mechanotransduction (Paluch *et al.*, 2015). These interactions between fibroblasts and the ECM they produce is of key interest in understanding emergent patterning in the stroma.

Research Question 2: What is the effect of the mechanistic interplay between CAFs and ECM?

VEGF signalling by CAFs induces the growth of new blood vessels (angiogenesis) towards and eventually into the tumour, supplying the tumour with the resources it needs to thrive and also as a means for metastasis (Bentley, Gerhardt and Bates, 2008; Bentley *et al.*, 2014). For cancer to successfully metastasize, first cancer cells need to be able to break away from the main tumour into the blood stream. They then need to travel through the blood stream, leave the blood stream and colonise a new location. It seems almost impossible for a cell to be able to accomplish this journey without dying and yet it is metastasis, and not the primary tumour, which kills the vast majority of cancer patients. This troubling and fascinating problem needs to be far better explored and may be an area in which probabilistic mathematical modelling could be hugely insightful. In all stages of tumour development, the stroma plays an essential role, helping the cancer cells get nutrients, migrate and ultimately metastasise (Valastyan and Weinberg, 2011).

1.3 Fibroblasts and cancer-associated fibroblasts

Fibroblasts are an abundant cell-type throughout the body, typically found individually in interstitial space in a quiescent state (Kalluri, 2016). Fibroblasts become activated commonly through PDGF signalling during wound healing (Lynch *et al.*, 1987). They become more motile, travelling to the wound site, proliferate more and produce and organise the ECM at the wound, which becomes the scar tissue (McDougall *et al.*, 2006). After the wound is healed, the fibroblasts typically apoptose or return to their inactive state (Basan *et al.*, 2013). In cancer, fibroblasts remain activated and so cancer has often been described as “a wound that does not heal” (Karagiannis *et al.*, 2012). These activated fibroblasts have properties which lend themselves to ECM generation and organisation such as increased contractility, proliferation and proteolytic capabilities. Activated fibroblasts also gain expression of α-smooth muscle actin (αSMA), which is often used as a biological marker for CAFs.

CAFs are recruited and reprogrammed by cytokine signalling by the tumour and metastatic sites (LeBleu and Kalluri, 2018) and are often found at the tumour

periphery (Egeblad *et al.*, 2008). CAFs can derive from a number of origins such as resident fibroblasts, bone marrow or resident pericytes (Augsten *et al.*, 2014; Karagiannis *et al.*, 2014; LeBleu and Kalluri, 2018). CAFs are highly heterogeneous due to their source and location in the TME. Environmental conditions such as spatial confinement or lack of nutrients can lead to CAFs modifying their phenotype. This plasticity of CAFs makes them often hard to characterise.
In their activated state, CAFs share many properties of activated fibroblasts in wound healing, producing high levels of matrix metalloproteinases (MMPs) for ECM reorganisation and secreting other signals crucial for tumour development (Feig *et al.*, 2013; Kalluri, 2016). Whilst there has been some indication that certain components of the CAF secretome might be tumour inhibitory (Özdemir *et al.*, 2014), the vast majority of literature indicates that CAFs promote tumour progression (LeBleu and Kalluri, 2018). This is done through the remodelling of the ECM to facilitate tumour invasion and the secretion of various signalling molecules such as VEGF (promoting angiogenesis), TGFβ and PDGF (promoting cell division and the recruitment of more CAFs to the tumour). In these ways, CAFs and consequently the ECM, play a crucial role in tumour progression and response to therapies.

Different spatial patterns of ECM have previously been observed in culture (Elsdale, 1968) and *in vivo* (Figure 2). Fibroblasts derived from human and mouse tumours display different emerging patterns *in vitro* as cells move from low density to confluence. Certain CAF populations can organise their cell bodies so that they align in terms of orientation and direction of movement. Other CAF populations remain largely non-aligned. Importantly, the collective behaviour of alignment through organisation of cell bodies corresponds to the organisation of underlying ECM. Aligned ECM is associated with disease progression and it has been shown that cancer cells have increased migratory capabilities when moving on aligned, isotropic matrices (Provenzano *et al.*, 2006; Goetz *et al.*, 2011; Drifka *et al.*, 2016a). This is illustrated schematically in Figure 3.

Additionally, work by the groups of Keely and Eliceiri (Provenzano *et al.*, 2006, 2008; Drifka *et al.*, 2016b) has shown that aligned collagen leads to cancer

progression. They define three Tumour-Associated Collagen Signatures (TACS) for characterising collagen organisation:

- TACS1 – increase collagen density around tumours as compared to normal tissue
- TACS2 – straightened collagen fibres running tangential to the tumour.
- TACS3 – straightened collagen fibres oriented radially to the tumour mass.

All of these signatures are associated with increased cancer invasion and tumours typically progress from TACS1 through to TACS3, although this conversion process is currently not understood. Crucially, it is the aligned organisation and increased density of collagen which have been shown to promote cancer invasion. Both of these properties are associated with CAFs.

In order to reduce the invasive potential of a tumour, it is paramount to understand the mechanisms causing CAF alignment. Differences in cell motility mechanisms are likely to underlie this variability collective cell behaviour and by implication, ECM organisation.

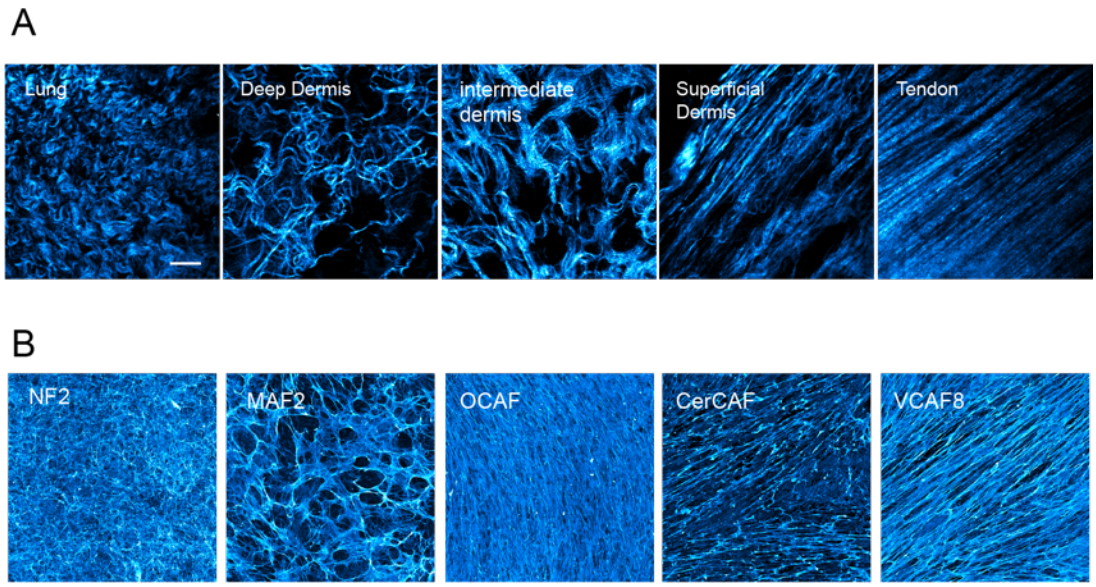


Figure 2: ECM patterns

(A) *In vivo* diversity of collagen organisation in healthy tissue in mice by second harmonic imaging, a microscopy technique suitable for imaging collagen fibres (scale bar 20μm). (B) *In vitro* images of matrix produced by normal fibroblasts (NF2, left) and different CAF populations.

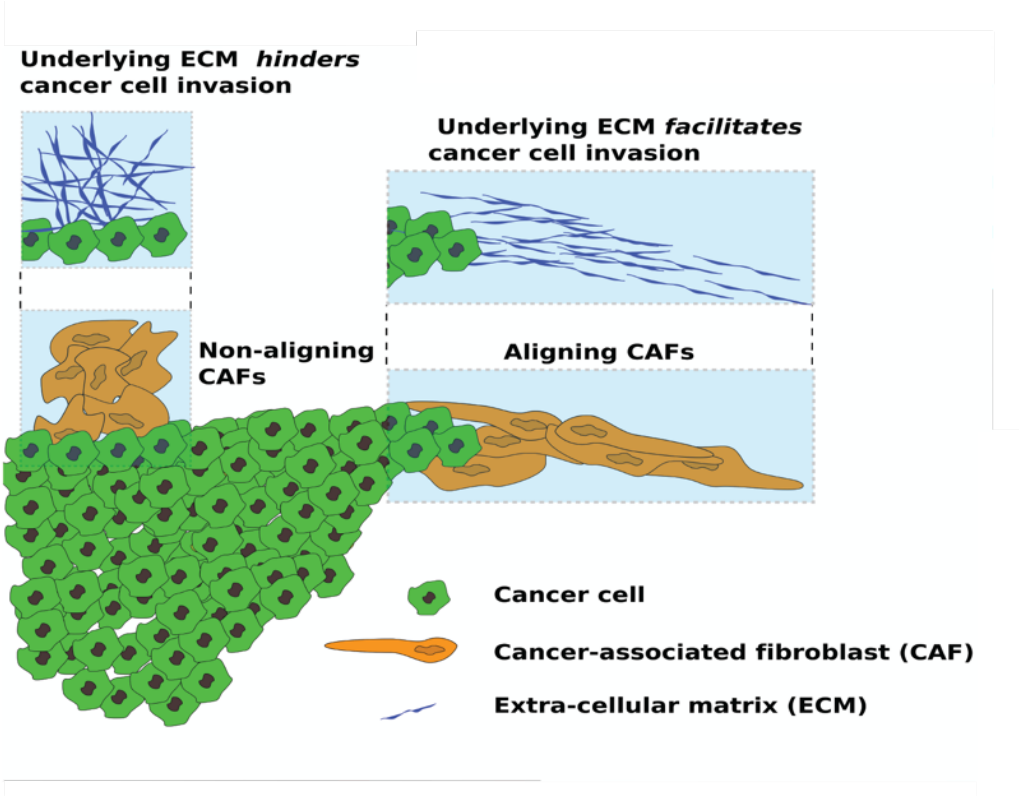


Figure 3: Schematic of CAF spatial organisation in TME

Two different CAF populations (orange) adjacent to the cancer cells (green) produce different ECM patterning (blue). Aligned ECM (right population) has been shown to facilitate cancer invasion by acting as tracks along which the cancer cells can more easily migrate.

1.4 Rules governing cell motility mechanisms and behaviour

1.4.1 Cell motility

In order for a cell to successfully migrate, a sequence of steps needs to be followed: firstly, actin-rich protrusions are generated at the edge of the cell. These protrusions cause the cells to have an asymmetric morphology with a leading edge and trailing edge. Next, adhesion of these protrusions to the substrate occur at the front of the cell coupled with contraction of the cell body at the trailing edge. Finally, the cell is pulled forward by contractile forces generated largely by the actomyosin network (Ananthakrishnan and Ehrlicher, 2007; Petrie, Doyle and Yamada, 2009). Cells can preferentially form protrusions in response to directional external cues, for example in response to a chemical gradient (chemotaxis), ECM adhesion gradient (haptotaxis) or ECM stiffness gradient (durotaxis). In the absence of an external gradient, where there is either a uniform external cue or none at all, cells migrate according to their intrinsic directionality. The ability of a cell to form stable protrusions at a leading edge will determine how effectively the cell can migrate. The formation of protrusions and the contraction of the trailing edge are underpinned by the distribution and concentration of the Rho GTPases within the cell: RhoA, Rac1 and Cdc42 (Mayor and Carmona-Fontaine, 2010; Marée, Grieneisen and Edelstein-Keshet, 2012). Cells with distinct and stable distributions of these GTPases are highly polarised which manifests as the cell having a clear leading edge with cell protrusions, a trailing edge, and migrational persistence (Petrie, Doyle and Yamada, 2009). In this way cell polarity and migrational persistence are coupled. In this work migrational persistence is used as a measure of CAF polarity.

The polarity mechanisms by which cells are able to respond to intrinsic and extrinsic cues are nuanced. A cell must be highly sensitive to small asymmetries in external cues and remain so even once large protrusions have been formed, thus allowing a cell to repolarise in a different direction. Meinhardt (Meinhardt, 1999) describes an elegant minimal model which can account for such behaviour by means of an oscillating inhibitor, periodically preventing sensitivity to an extrinsic cue. By varying the range and period of the inhibitor, a range of observable polarity

behaviours are observed such as multiple protrusions, large protrusions and the formation of protrusions through intrinsic cues alone.

1.4.2 Cell-cell interactions

CAFs, like many other cells, can transfer information locally between neighbours. This transfer of information could be via N-cadherin mediated signalling (Paluch *et al.*, 2015), force transmission through cell-cell adhesions similar to the phenomenon of plithotaxis observed in epithelial sheets (Trepat and Fredberg, 2011; Zaritsky *et al.*, 2015), mechanotaxis, cell-cell signalling or *via* communication between CAFs and the ECM which they create and maintain (Shiga *et al.*, 2015; Li *et al.*, 2017). Competing with these local cooperative mechanisms is the inability for two cells to occupy the same space and the frequently observed repolarisation of migration upon contact, termed contact inhibition of locomotion (CIL) (Maiuri *et al.*, 2015; Roycroft and Mayor, 2015; Mayor and Etienne-Manneville, 2016; Stramer and Mayor, 2016; Zimmermann *et al.*, 2016). It has been shown that CIL response is largely dependent on collision geometry (Desai *et al.*, 2013). Any exploration of collective cell behaviour will have to carefully consider these complex dynamics.

1.5 Spatial organisation of the tumour

The commonality of the stroma to all cancers makes it an attractive area to try to understand better as a means of inhibiting cancer progression. In particular, there is currently a lot of research being done to better understand the spatial structure of the stroma in conjunction with the temporal dynamics of the tumour (Egeblad *et al.*, 2008; Heindl, Nawaz and Yuan, 2015; Leung, Rice and Barton, 2015).

Studying the architecture of the stroma has two obvious potential consequences: classification and prediction. It seems likely that in the near future it will be possible to derive the stromal structure of large areas of the tumour in patients (Heindl, Nawaz and Yuan, 2015). This would allow for patient stratification, for instance, a patient with a large proportion of CAFs may have a poorer prognosis than a patient with a large proportion of T-cells, leading to different therapies.

Further, building a spatial map of the TME could allow for identification of high-risk metastatic points within the tumour and the potential to manipulate these areas. For example, it has been shown that the altered metabolism of cancer cells can create gradients of ECM metabolites which orchestrate phenotypic diversity in the TME and the differentiation of cells into distinct subpopulations (Carmona-Fontaine *et al.*, 2017). Another example, is that areas of aligned ECM could be indicative of where invasion is likely to occur (Provenzano *et al.*, 2006). Understanding the TME architecture may even enable the prediction of cancer cell invasion. As previously documented (Friedl *et al.*, 2012), cancer cells are able to invade surrounding tissue as single cells, in multicellular streams or as part of coordinated collectives involving many cells. The mode of invasion is thought to be dependent on cell-cell adhesions, cell contractility and cell-matrix adhesions. It is likely that stromal organisation will also be an important determinant. These are just a few examples of potential benefits of understanding the spatial organisation of the TME.

Research Question 3: Can spatial organisation of matrix predict patient outcome?

The importance and variation of spatial structure in tumours leads to natural parallels with ecology (Maley *et al.*, 2017). Indeed, in ecology the most effective way to eliminate a species is by destroying its niche environment (Yuan, 2016). In the same way, therapies which target the specific TME can be an effective means to treat cancer. In some instances, organisation of the TME is thought to be a possible cause of drug resistance. One example of the importance of spatial organisation in tumours can be found in the diversity of vasculature; from well organised to more tortuous vessels (Manning *et al.*, 2013) associated with hypoxia and metastasis, together with radiation responses (Scott *et al.*, 2016). A recent study has found that hypoxia causes an increase in collagen I and LOX production in ovarian cancer and that inhibiting LOX (and thereby reducing crosslinking) reduces tumour burden in mice (Natarajan *et al.*, 2019). Directly manipulating the ECM therefore has potential therapeutic benefits. Another example is using the location of immune cells in the TME as a prognostic factor in colorectal cancer (Nawaz *et al.*, 2015; Yuan, 2016).

The Bates group has a long history of agent-based modelling at the cellular/subcellular level including unravelling how angiogenic sprout formation occurs. Prior work has investigated different modes of cancer cell migration through different ECM geometries with a finite element model (Tozluoğlu *et al.*, 2013a). A cell has an actomyosin cortex and an outer cell membrane, both modelled by a set of agents. The model combines actin-polymerisation, contractility, blebbing and cell-ECM adhesion and is able to established requirements for effective migration in different ECM geometries. The model is robustly able to predict experimental drug perturbations and is a good example of how agent-based modelling can be a powerful tool for teasing apart complexity in biological systems.

Thinking of tumours in broad spatial terms has recently attracted renewed focus. Much work has been done to understand the temporal evolution of tumours however, in practice it is difficult to get temporal information, particularly in patients. Spatial information is much more readily available through imaging of multiple biopsy sites, therefore using “space as a surrogate” for time (Yuan, 2016).

The TME is an ecosystem in which different entities such as cancer cells, CAFs, immune cells and ECM coexist and influence one another. Metrics such as colocalisation between “species”, clustering, roughness of the interface between species and alignment can provide important information as to how a tumour is likely to evolve both spatially and temporally.

Spatial metrics have been shown in some instances to be more effective prognostic markers than traditional histopathology methods alone (tumour size, tumour grade etc.). It is certain that with the availability of tumour images and the advancements in machine learning techniques allowing for processing of these images, spatial organisation of the tumour will become increasingly important as a prognostic tool.

Methods for studying the spatial organisation of tumours are developing rapidly. Machine learning is one such method, principally focussed on identifying patterns, and thus is of crucial importance in the study of TME architecture. In particular, a Convolutional Neural Network (CNN) is a class of neural network optimal for analysing patterns in visual data. CNNs fit into the sub-family of machine learning

called deep learning (Lecun, Bengio and Hinton, 2015), which uses multiple layers of analysis to extract increasingly abstract patterns. There is currently huge excitement surrounding deep learning, and important advances have been made taking advantage of the huge amount of data now available. Some examples of this are in text mining, language interpretation and driverless cars. CNNs are particularly useful in analysing images because they can recognise spatial patterns.

Weaknesses of CNNs (and neural networks in general) are that the computational cost is high and a large amount of training data is typically required to generate meaningful results. However, these obstacles are not insurmountable, with GPUs becoming more widely used, vastly increasing computational efficiency. It is also clear that when enough training data can be obtained from “expert opinion”, CNNs can produce impressive results. Applications of CNNs to the field of tumour architecture have included using H&E stained samples and expert opinion to train a classifier for different cell types (Bulten *et al.*, 2019). The main gap in deep learning is in the “black box” effect, whereby it is often hard to understand why the neural network has chosen to classify images according to various abstract properties. Whilst CNNS can effectively identify patterns it is often unclear how exactly they form and operate. This can lead to a lack in understanding of how phenomena arise.

As discussed in (Baker *et al.*, 2018), it is important to consider how these emerging machine learning techniques and traditional mathematical modelling fit together. Broadly, mathematical modelling can be used to study cause, whilst machine learning can be used to study effect and patterns. Individual cells and matrix fibres are endowed with certain properties that enable them to interact and cause different collective behaviours and heterogeneous spatial patterning. This resulting architecture is the effect of those properties. Machine learning has the capability to identify patterns across huge datasets (outputs), which is clearly of great use in the age of big data and data abundance. Simultaneously, mathematical modelling remains indispensable for understanding the fundamental principles governing the relationship between inputs (cell properties and rules) and outputs (matrix patterns) (Baker *et al.*, 2018). Therefore, in the context of understanding TME architecture, mathematical modelling and machine learning can be used in synergy to increase

understanding. In this vein, whilst the focus of this work is on mathematical modelling, complementary machine learning approaches are explored in each chapter.

1.6 Modelling approach

For many decades now, mathematical modelling has been viewed as an essential tool for understanding biological phenomena. Models can provide mechanistic understanding of how complex behaviour arises from a set framework, extrapolate and make predictions about how systems will evolve and provide an essential perspective in tackling biological problems. Modelling is often at its most powerful when combined in synergy with experimentation. A feedback loop between computational and experimental scientists enables more meaningful insights. By incrementally building and improving models, with assumptions that are confirmed with experimentation, predictions can be made which can in turn direct the course of experiments. Mathematical models are fundamental in reducing complex problems down to their key components and developing a core understanding of how such systems function (Mogilner, Allard and Wollman, 2012).

Many types of models have been used to study cell behaviour, from single cell models, to continuum modelling of entire organs. An overview of different collective cell modelling approaches is given here. Whilst this summary of models is by no means exhaustive, it aims to give a snapshot into common model choices and endeavours to justify why the framework used in the rest of this thesis is an appropriate choice.

1.6.1 Which dimension to work in?

The majority of models looking at collective cell behaviour are in 2D. This is because the majority of experimental work in this area takes place *in vitro* in a 2D environment and because of computational limitations. However, cells can move in a very different way in 3D with basal and apical constraints (and therefore *in vivo*) and so this is an important future direction to consider moving into (Petrie, Doyle

and Yamada, 2009; P.-H. Wu *et al.*, 2014; Steinwachs *et al.*, 2015; Wang *et al.*, 2016).

In 2D, cell migration is largely achieved by actin polymerisation at the front of the cell, producing stable filopodia. These filopodia adhere to the substrate, whilst the rear of the cell de-adheres from the substrate (Ananthakrishnan and Ehrlicher, 2007). Through the generation of contractile forces, the cell is then able to crawl forwards. This process of cell motility can be modelled as a persistent random walk (PRW) (P.-H. Wu *et al.*, 2014). In 3D (and 1D) however, cells can have apical and lateral confinement in addition to the basal confinement experienced in 2D (Pizzo *et al.*, 2005; Sandersius and Newman, 2008). These additional constraints cause the cell to migrate in a more bleb-like manner, producing exploratory pseudopodia and a more spindle-like shape as the cell tries to navigate through a complex 3D environment. It has been shown that cell movement in 3D does not follow a persistent random walk, exerts forces in different ways and that the cells have different phenotypes due to their different shapes.

1.7 Collective cell models

1.7.1 Continuum models

Continuum models have been able to offer profound insights into the tumour population as a whole, in particular applying populations dynamics and game theory methods (Swanson *et al.*, 2003; Matzavinos, 2004; Bearer *et al.*, 2009; A. Wu *et al.*, 2014; Notbohm *et al.*, 2016; Sartakhti *et al.*, 2016). The limitations of such models lie in the difficulty of assessing collective behaviour and individual cell interactions.

Work by Anderson et al. (Anderson *et al.*, 2007) considers a system of partial differential equations across three variables: tumour cell density, ECM density and matrix degrading enzymes, to study tumour invasion. The authors solve these PDEs in one and two dimensions to establish an analytical understanding of the interactions between the three variables during invasion. They then use these findings as a basis for developing a discrete model.

Another continuous model presented in (Escaff *et al.*, 2018) is specifically designed to consider flocking behaviours. Self-propelled particles move with a continuous-time persistent random walk in one-dimension, and particles are able to influence the motility of each other. The model has two main parameters: interaction range and interaction strength. By varying these two parameters the authors map a phase diagram showing transitions between disordered, ordered homogeneous and ordered clustering behaviours. This is an elegant model which can be solved analytically, but becomes complicated even when limited to the one-dimensional case.

Continuum models enable the exploration of analytical solutions but can lack detail. On the other hand, discrete models are more adaptable but harder to analyse, depending on the level of detail. The complexity and multi-scale nature of cell interactions, involving feedback at the protein, cellular and tissue levels lead to a trade-off between complexity and scale (Lecaudey and Gilmour, 2006; Ellery *et al.*, 2016; Camley and Rappel, 2017).

1.7.2 Sub-Cellular Element model

The Sub-Cellular Element Model (SEM) (Newman, 2005; Sandersius, Weijer and Newman, 2011) is a creative model ideally used for studying cell rheology but with extensions to multicellular systems. Elements of the cell are initialised and move with respect to each other according to predefined Morse potentials. By defining different intra and inter-cellular potentials, it is possible to generate sophisticated cell shapes and interactions between groups of cells. This framework is also highly suitable for three-dimensional analysis. By using a Voxel-like approach, only interactions between elements that are sufficiently close to each other are computed, saving on computational cost. However, the SEM does not lend itself to studying the collective motion of cells, particularly at the scale of hundreds or thousands of cells. Further, how to define appropriate Morse potentials between elements is often ambiguous.

1.7.3 Cellular Potts model

One of the most important models for studying collective cell motions is the Cellular Potts Model (CPM). To list just a few examples, the CPM has been used to investigate swirling of cells (Chang *et al.*, 2013), cell morphology during gliding (Albert and Schwarz, 2016) and polarity (Marée, Grieneisen and Edelstein-Keshet, 2012). Originally introduced in a seminal paper in 1992 by Graner and Glazier to describe cell sorting, this lattice-based model allocates lattice sites to particular cells based on the minimisation of an energy function. Specifically, cell 1 occupies the region where lattice sites have "spin" equal to 1. Lattice sites not allocated to any spin are considered to be surrounding media. A system's energy is defined by the Hamiltonian

$$H = \sum_{\substack{\text{neighbouring} \\ \text{sites } a,b}} J_{ab}(1 - \delta(\sigma_a, \sigma_b)) + \lambda \sum_{\substack{\text{cells} \\ i}} \left(A_i - A_{i,0}\right)^2 \quad \textbf{(1)}$$

Where J_{ab} is the energy between neighbouring sites $a \text{ and } b$. The δ term is a Dirac function such that if two neighbouring sites have the same spin (value) then there is no energy cost. The second term describes a penalty (λ) on cells of area A_i for deviating away from their target area ($A_{i,0}$). The model evolves by attempting to minimise the Hamiltonian through Monte Carlo steps.

The model is easily extendable by adding terms to the Hamiltonian. The CPM is computationally inexpensive and easily repurposed to study new phenomenon, making it widely used in modelling single and collective cell motility (Scianna and Preziosi, 2013; Tozluoğlu *et al.*, 2013a; Osborne *et al.*, 2017).

One minor criticism of the CPM is that cell motility is inextricably linked with changes in the cell boundary (Camley and Rappel, 2017) which may not be realistic. However, in the context of answering the key questions posed in this thesis, the primary problem of the CPM is in modelling cells and ECM fibres simultaneously. This has been attempted (Scianna and Preziosi, 2013), but it is difficult with the CPM to model explicit fibres that behave in a physiologically

reasonably way and to implement rules for interaction specifically between the cells and individual fibres without making the system far more complicated.

1.7.4 Vertex/Voronoi models

Two other well-established models of collective cell behaviour are the closely related Vertex and Voronoi models (Sánchez-Gutiérrez *et al.*, 2015; Camley and Rappel, 2017), whereby cells are defined with respect to their boundaries. These systems evolve once again by an energy minimisation task, but the Hamiltonian refers specifically to cell area and perimeter. Thus, the Hamiltonian is described as

$$H = \sum_{\substack{\text{cells} \\ i}} [K_{area}(A_i - A_0)^2 + K_{perim}(P_i - P_0)^2] \quad \textbf{(2)}$$

The Voronoi model varies from the Vertex model in that equations are solved for the centre of mass on each cell as opposed to solving the energy equation for each of the specified vertices. Vertices are not specifically defined, rather the region of the cell is allocated by Voronoi tessellation. This feature gives the Voronoi model the advantage of being able to handle cell rearrangements more easily than the Vertex model, but it can also be challenging to define the free boundary problem (Camley and Rappel, 2017). Both models are ideally suited to studying confluent layers of cells, but do not comfortably adapt sub-confluent systems. An interesting study applying a Vertex model, Voronoi model and CPM to the same biological problems elucidates the differences between the models (Osborne *et al.*, 2017). Their findings point to reasonable qualitative agreement between the models for a variety of problems, but that one model is often more suitable than the others for a specific problem.

1.7.5 Force-based models

There have been a number of insightful force-based models investigating chemoattraction and CIL in recent years (Woods *et al.*, 2014; Camley *et al.*, 2016; Szabó and Mayor, 2016). Based on simple principles of CIL and co-attraction, cells can migrate collectively. However, it is often challenging to model the complexity of cellular forces and measure them accurately. Trepat et al. (Trepat *et al.*, 2009) have shown that cells engage in a global "tug of war", pulling in many different directions simultaneously. Several alignment mechanisms in collective motion have been proposed in the literature (Camley and Rappel, 2017), which circumvent the challenges of measuring forces: velocity alignment where a cell's polarity aligns to its velocity (Basan *et al.*, 2013; Camley *et al.*, 2014) and neighbour alignment (Szabó *et al.*, 2006; Chaté *et al.*, 2008; Sepúlveda *et al.*, 2013). So far, these models have shown good agreement with experimental work and the justification for these mechanisms is being explored through force measurements. For instance, the phenomenon of plithotaxis has been observed in epithelia where there is a tendency for cells to align their direction of migration with the axis of maximal principal stress (Trepat and Fredberg, 2011; Zaritsky *et al.*, 2015). A major class of model using neighbour alignment is the Vicsek model, which is detailed in Section 1.7.7.

1.7.6 A hybrid model

An alternative to choosing between continuum and discrete models of collective behaviour is the hybrid agent-based model developed by the Maini group (Dallon, Sherratt and Maini, 1999; McDougall *et al.*, 2006), in which cells are modelled as discrete agents and the underlying ECM is modelled as a continuum. A brief description of the model is given here. The path of cell i denoted by $\boldsymbol{f}^i(t)$ and the ECM at position $\boldsymbol{x} = (x, y)$ is denoted $c(\boldsymbol{x}, t)$. Cells receive directional cues from the matrix and move in the direction of the weighted average of their persistence and the matrix guidance. Fibroblasts deposit ECM in the same orientation as their direction of movement. Further, fibre deposition and degradation rates are assumed to be equal so collagen density is constant. Matrix orientation changes as a weighted average of the orientations of nearby fibroblasts. The authors use their

model to investigate ECM pattern formation in particular within the context of wound healing and matrix remodelling. They find that for all but low values of cell migrational persistence, cells will deviate from a straight path, producing a more swirl-like ECM. Additionally, Dallon et al. find that seeding a strip of aligned fibres that is sufficiently thick causes the rest of the matrix to align similarly. The work also explores modelling fibrin and collagen explicitly in wound healing. This model is of great interest, but does not include any modelling of cell-cell interactions. Further, modelling the ECM as a continuum prevents direct interaction between specific fibres and the fibroblasts.

1.7.7 Vicsek model

As documented above, there are many useful and well-established models for collective behaviour. These models, as well as plethora of others have been adapted to address many different problems in different systems. However, this can often lead to over-parameterised models. The seminal work by Vicsek et al. (Vicsek *et al.*, 1995) in 1995 aimed to strip back these models to identify the simplest model that could generate collective flocking behaviour in diverse systems (Battersby, 2015). The Vicsek model is generalisable, tractable and displays remarkable emergent behaviour given its simplicity.

The framework of the Vicsek model is as follows: At initialisation particles are seeded with random orientation in a two-dimensional box with side length L and periodic boundaries. Particles have no area and all particles move at the same fixed speed s. At every time step particle orientations and then particle positions are simultaneously updated. The position of particle i at time $t + \Delta t$ is given by

$$\underline{x}_i(t + \Delta t) = \underline{x}_i(t) + \underline{v}_i(t + \Delta t)\Delta t \quad \textbf{(3)}$$

where

$$\underline{v}_i(t + \Delta t) = s \begin{pmatrix} \cos(\theta_i(t + \Delta t)) \\ \sin(\theta_i(t + \Delta t)) \end{pmatrix} \quad \textbf{(4)}$$

With s representing the fixed cell speed of all particles and

$$\theta_i(t+\Delta t) = \langle \theta_j(t) \rangle_{|\underline{x}_i - \underline{x}_j| < r} + \eta_i(t), \quad \textbf{(5)}$$

where $\langle \theta_j(t) \rangle_{|\underline{x}_i - \underline{x}_j| < r}$ is computed as

$$\langle \theta_j(t) \rangle_{|\underline{x}_i - \underline{x}_j| < r} = \tan^{-1}\left(\frac{\langle \sin\left(\theta_j(t)\right) \rangle_{|\underline{x}_i - \underline{x}_j| < r}}{\langle \cos\left(\theta_j(t)\right) \rangle_{|\underline{x}_i - \underline{x}_j| < r}} \right). \quad \textbf{(6)}$$

The $\eta_i(t)$ term represents the intrinsic noise of the particle, and is chosen at each time step from the distribution $\eta \sim U\left[-\frac{\zeta}{2}, \frac{\zeta}{2}\right]$.

The model parameters are neighbourhood size (r), and intrinsic noise (η) and particle density $\left(\frac{L}{N}\right)$, where L denotes the side-lengths of the square in which simulations take place and N denotes the number of particles. Particles with large neighbourhoods are well able to coordinate themselves with their neighbours. However, if particle density is too low, particles cannot coordinate. The noise term can be thought of as the "free will" of each particle or its adversity to cooperation. By varying these three parameters, discontinuous phase transitions are observed between order and disorder. Thus, the behaviour of a particle within the Vicsek model is determined by three components: an awareness of one's neighbours, a willingness to cooperate and enough neighbours to cooperate with. These fundamental ingredients are common to all collective systems. Importantly, diverse collective behaviour can be obtained without the requirement for population heterogeneity or leader/follower cells (Grégoire, Chaté and Tu, 2003).

The cell speed s has been described in the literature as a "thermalisation parameter" (Baglietto, Albano and Candia, 2012) in Vicsek model, relating to how long two particles will remain with a distance r from each other and be considered neighbours. Baglietto et al. found that in instances of low thermalisation, where particles move very quickly, the system will be more disordered. However, in instances of medium and high thermalisation when particles move slower, changes in s has a minimal effect on the system and does not affect phase transitions. Similarly, maintaining particle speed but taking a very large time step will cause

particles to spend less time as neighbours, altering the emergent behaviour. However, reasonable choices of time step and particle speed have been shown to have little impact on the model outcome (Dallon, Sherratt and Maini, 1999; Baglietto, Albano and Candia, 2012).

Extensions to the Vicsek model include considering nematic particles (Ginelli *et al.*, 2010) and a surrounding fluid (Chaté *et al.*, 2008) and into three dimensions with zones of attraction and repulsion (Couzin *et al.*, 2002; Grégoire, Chaté and Tu, 2003). An extension of the Vicsek model described by Schumacher et al. (Schumacher, Maini and Baker, 2017) describes a term for flocking corresponding to averaging of neighbour orientations, a noise component and an additional term for attraction-repulsion forces between cells in three dimensions, as described in (Grégoire, Chaté and Tu, 2003). They use this model to show that seemingly heterogeneous behaviour can arise from the collective behaviour of homogeneous cells. The minimality and generalisability of the Vicsek model make it an attractive framework to study collective behaviour, which can be simply and naturally adapted to specific systems.

1.7.8 Modelling the effects of cell biochemistry on collective behaviours

One of the main challenges in mathematical modelling of collective cell behaviour is decoupling the mechanical and chemical properties of the system. Unlike inactive matter such as liquid crystals (Marchetti *et al.*, 2013), cells undergo orchestrated chemical responses to chemoattractants; therefore, cell-cell contacts and have intrinsically heterogeneous biochemical properties (Davis *et al.*, 2015).

Chemoattractants are common in biological systems such as wound healing, where a chemical signalling gradient originating at the site of the wound causes cells to preferentially migrate in the direction of the wound. A chemoattractant alters cell motility and therefore impacts upon collective cell behaviour. A significant body of work has explored modelling of chemoattractants, including work by De Palo *et. al* (De Palo, Yi and Endres, 2017), which combines a variant of the Vicsek model with a chemoattractant. The authors study how collective behaviour of Dictyostelium

arises as a result of cAMP signalling (a chemoattractant). In their model, the concentration of cAMP that a cell senses is given by the sum of cAMP secretion of all neighbouring cells.

Heterogeneity amongst cells can change the resulting emergent behaviour. It is well documented that CAFs are highly heterogeneous (Augsten, 2014). Systems incorporating the coexistence of several cell types can produce interesting chase and run behaviour (Osborne *et al.*, 2017). Heterogeneity at the single cell level can cause cells to become “leaders”, with the other cells designated as “followers”. However, there is still some debate as to how clearly these leader/follower roles are defined (Grégoire, Chaté and Tu, 2003; Schumacher, Maini and Baker, 2017).

The additional biochemical properties of cells can vastly increase the complexity of the system. It is therefore important to consider not only how such properties might affect the emergent behaviour, but also to recognise the trade-off between biological realism and mathematical tractability as offered by the Vicsek model.

1.8 Choosing an appropriate model

As demonstrated in this overview, there are many diverse and creative models for studying collective behaviour. The choice of model depends largely on the scientist’s aims. In choosing a suitable model, some important factors must be taken into consideration:

- Scale of system
- Mathematical tractability
- Computational cost
- Natural fit

Further, there are some models that lend themselves to particular problems, for example Potts modelling for cell sorting (Graner and Glazier, 1992) and Vicsek modelling for collective motion (Vicsek *et al.*, 1995). This is what is meant here by a natural fit of a model to a specific problem.

Earlier in this chapter, three research questions were introduced:

Q1 What properties of CAFs cause different collective behaviours?
Q2 What is the effect of the mechanistic interplay between CAFs and ECM?
Q3 Can spatial organisation of matrix predict patient outcome?

In order to begin to answer these questions, a mathematical model to explore such questions must satisfy certain criteria:

C1 Capability to model thousands of cells (scale)
C2 Capability to understand the origins of emergent behaviour (tractability)
C3 Capability to model many parameter perturbations (cost)
C4 A clear way to model ECM and interactions with cells (natural fit).

Given the aims and requirements of the model in this work, the Vicsek model framework was chosen. Table 1 gives a summary of all the models introduced and how they satisfy the modelling criteria C1-C4, to justify why the Vicsek model was chosen.

It is worth considering the extent to which the number of spatial dimensions of the model impacts the criteria. It is clear that modelling in one dimension is able to capture a very restricted set of cellular interactions and does not recapitulate the *in vitro* experiments of alignment (failing C4: natural fit). Whilst modelling in three dimensions is becoming more prevalent with increased computational power, there is still a considerable computational cost (failing C1 and C3 for now) and a lack of experimental data to inform how a rule-based system of cells interacting in three dimensions should be defined (failing C2). A two-dimensional model is able to satisfy the criteria, with the caveat that the two-dimensional behaviour is not the full story and there is much work to do in the three dimensional domain.

	Section	Refs	Advantages	Disadvantages	Criteria met
Continuum	1.7.1	(Anderson *et al.* , 2007)	Numerical solutions, general overview	No individual cells, many assumptions	C1, C3
Subcellular element model	1.7.2	(Sandersius, Weijer and Newman, 2011)	Rheology, flexibility	Expensive, measuring forces tricky	C2
Cellular Potts model	1.7.3	(Graner and Glazier, 1992)	Adaptable, scalable, elegant	No natural way to include ECM fibres	C1, C2, C3
Vertex/Voronoi	1.7.4	(Sánchez-Gutiérrez *et al.* , 2015)	Well studied, good for cell sorting	Less sensible for sub-confluence	C1, C2
Force-based	1.7.5	(Woods *et al.* , 2014)	In-depth, natural fit	Difficult to measure force, hard to scale up	C4
Hybrid	1.7.6	(McDougall *et al.* , 2006)	Elegant, cheap, justifiable	No individual fibres	C1, C2, C3, C4
Vicsek	1.7.7	(Vicsek *et al.* , 1995)	Minimal, adaptable, tractable	Overly simplistic neighbour alignment	C1, C2, C3, C4

Table 1: Summary of models for collective behaviour

1.9 A note on experimental work

It is important to note that, unless otherwise indicated, all experimental work and resulting *in vitro* and *in vivo* images shown throughout this thesis were performed or obtained by Danielle Park from the Tumour Cell Biology Laboratory at the Francis Crick Institute. Experimental methods are given in Appendix for interest but were written by Danielle Park and not by the author.

Chapter 2. Modelling fibroblast interactions

2.1 Introduction

In addition to diverse ECM organisation observed *in vivo* (Provenzano *et al.*, 2008; Conklin *et al.*, 2011; Mayorca-Guiliani *et al.*, 2017; Park *et al.*, 2019), different spatial patterns have previously been observed in fibroblasts in culture. It is of paramount importance to understand how such higher-order organisation arises. A whole library of *in vitro* ECM was generated from different CAF and normal fibroblast populations, ranging from isotropic to highly anisotropic patterns (Park *et al.*, 2019). Two example CAF populations from either end of this alignment spectrum are studied here in order to explore the key differences leading to such divergent patterning behaviour. CAFs were derived from vulval and mammary carcinomas and observed transitioning from low to high confluence through timelapse imaging over seven days (Figure 4). Over time, one CAF population, called VCAF8, began to align in terms of orientation and direction of movement, whilst the other CAF population, called CAF1, did not. The fact that different ECM patterns are observed *in vitro* indicates that there are differences at the cellular level between CAF populations that facilitate higher-order patterning. Understanding what these mechanistic differences are, and how this affects emergent behaviour is, the subject of this chapter.

2.2 Motivated by experiments

2.2.1 Experimental observations

To determine whether these collective fibroblast behaviours were predictive of global ECM organisation, immunofluorescence staining of both aligning and non-aligning populations was undertaken. Indeed, fibronectin deposition corresponded with cell body orientation, indicated by F-actin. VCAF8 cells produced an aligned ECM, whilst CAF1 cells produced a disorganised meshwork (Figure 5). These two

CAF lines are for the remainder of this chapter referred to as “Aligning” and “Non-Aligning” respectively.

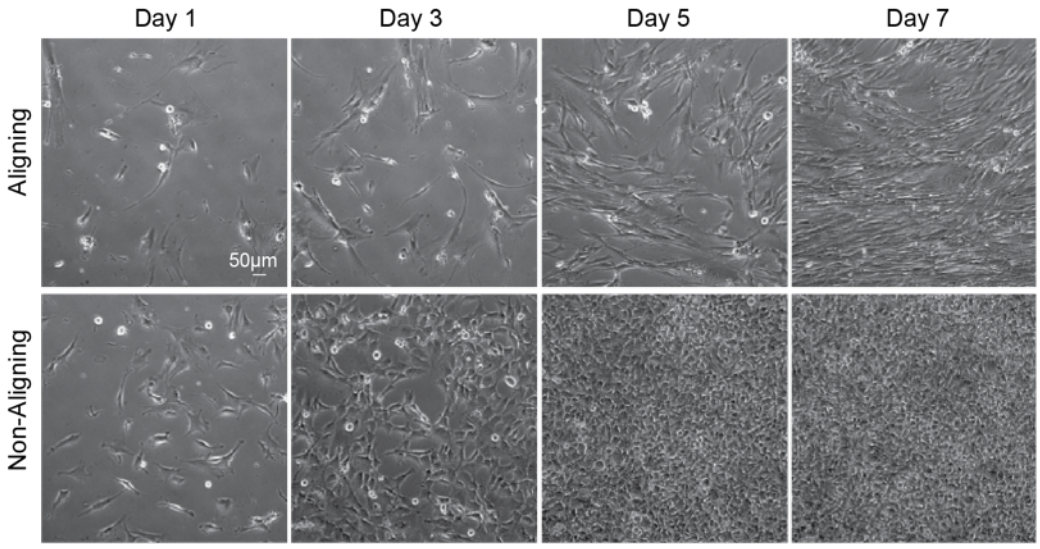


Figure 4: Timelapse phase contrast imaging of CAFs moving from sub-confluence to confluence over 7 days.

Certain fibroblast lines align from sub-confluence to confluence. Vulval carcinoma fibroblasts align cell bodies at confluence (upper panels) whilst breast carcinoma fibroblasts do not (bottom panels). Scale bar represents $50\mu m$.

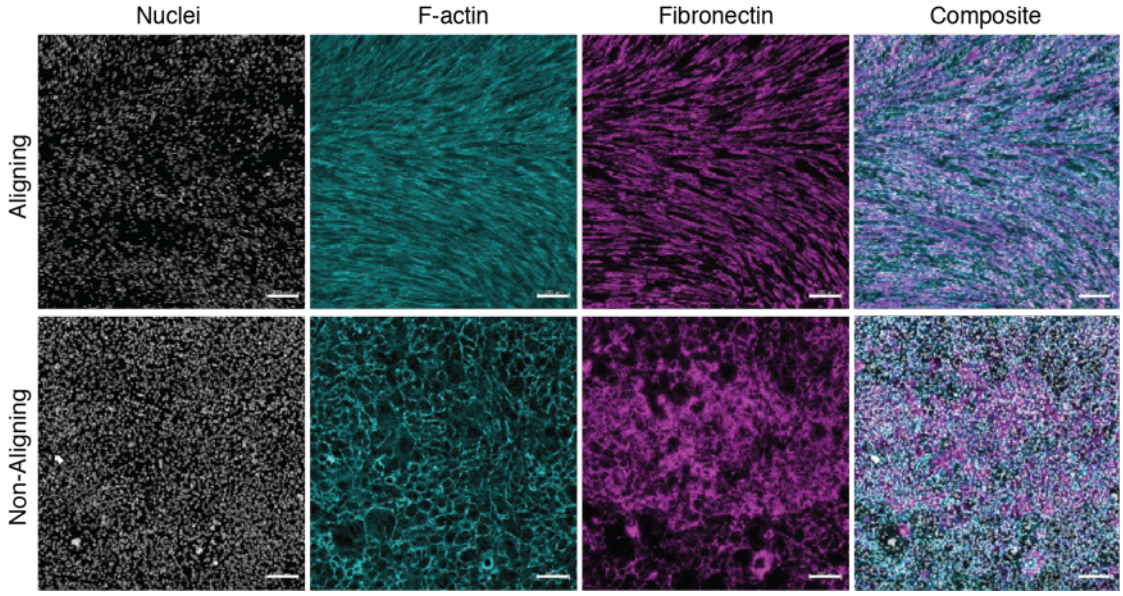


Figure 5: Immunofluorescence of aligning and non-aligning CAFs at confluence. From left to right: Image shows staining of the cell nuclei, cytoskeleton (F-actin) and underlying ECM (Fibronectin). Cell body organisation corresponds to ECM arrangement as shown in the right-most composite images. Scale bar represents $100\mu m$**.**

An important feature of the aligning CAFs is their nematicism. When CAFs align, they do not all move in the same direction, but rather some cells stream in the complete opposition direction forming antiparallel streams. When two CAFs align they do not appear to sense the polarity of each other, generating this nematic phenomenon (Marchetti *et al.*, 2013).

One key observation is that aligning CAFs exhibit alignment in sub-confluent conditions. Cells were nuclear labelled and tracked using the software Metamorph. These tracks for different days of the assay show cells aligning at early time points in the absence of spatial confinement. The lack of a requirement for high confluence and spatial confinement suggests that properties inherent to individual cells contribute to alignment.

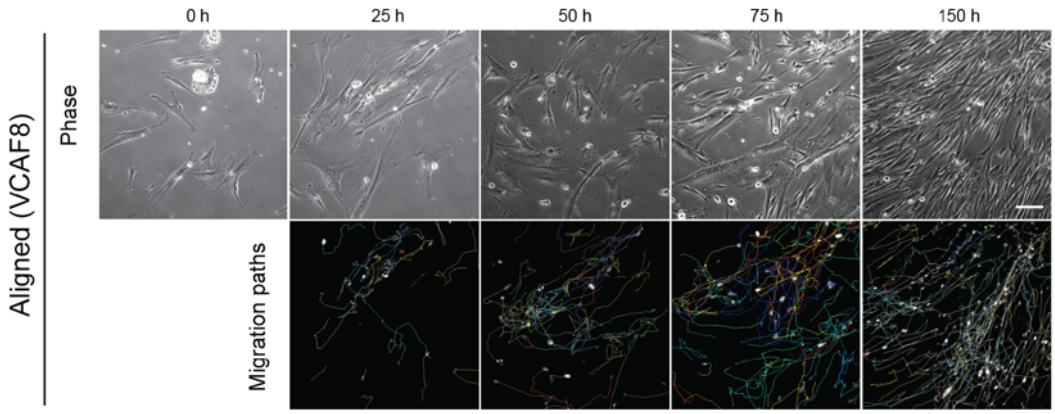


Figure 6: Tracking cell migration paths

Time-lapse microscopy of aligned (VCAF8) fibroblasts over seven days, showing bright field phase imaging (top panel) and nuclear tracking (bottom panel). The migration paths shown are of the previous 25 hrs. Cell body anisotropy occurs early at low confluence (25hrs) and is predictive of the final global organisation (150 hrs).

This gradual alignment over time was demonstrated quantitatively. A metric of alignment was developed that was able to quantify changes over time and space. Quantifying cell body alignment over multiple neighbourhood sizes captures alignment patterns at both local and global scales. At a specified time point t, for a cell i, for a given neighbourhood r between r_{min} and r_{max}, let the set J denote all cells j which have a distance in the range $[r_{min}, r_{max}]$ from i. In the experiments, distance is given in microns unless otherwise stated. The orientation of cell i denoted $\theta_i(t)$ is defined as the orientation of the line of best fit through the positional coordinates of cell i recorded at times $[t - 3\Delta t, t - 2\Delta t, t - \Delta t, t]$. The vector $\Delta\Theta_i$ is then defined as the antiparallel difference in angle between the orientation of cell i and the cells in J. The un-normalised angle of alignment of cell i at time t over a given distance $r = [r_{min}, r_{max}]$ can be defined as

$$u_{r,i}(t) = median(\Delta\Theta_i). \quad \textbf{(7)}$$

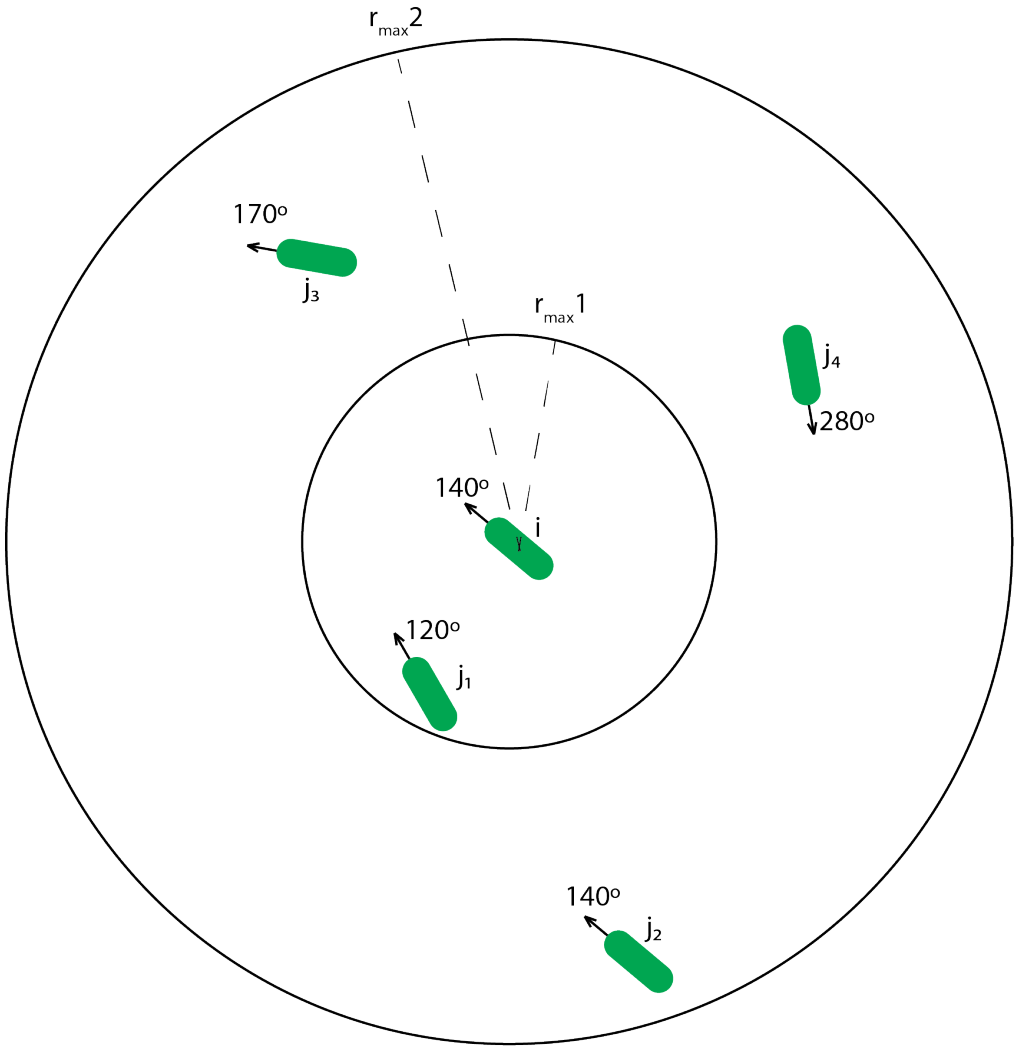
The value $u_{r,i}(t)$ returns a value of NA if the set J has fewer than five elements ie if cell i has fewer than five neighbours in the neighbourhood r. In the case where $r_{min} = 0$, the set J contains the cell i itself. The average un-normalised angle of alignment across all cells is given as

$$u_r(t) = \frac{1}{N}\sum_{i=1}^{N} u_{r,i}(t). \quad \textbf{(8)}$$

This can be normalised to give values between 0 and 1 as

$$a_r(t) = \max\left(0, 1 - \frac{u_r(t)}{\frac{\pi}{4}}\right). \quad \textbf{(9)}$$

A schematic showing the derivation of this metric is given in Figure 7.



$r_{min} = 0, \quad r_{max} = r_{max}1$

$J = (i, j_1), \quad n = 2$, too few neighbours

$r_{min} = 0, \quad r_{max} = r_{max}2$

$J = (i, j_1, j_2, j_3, j_4), \quad n = 5$

$= (0°, 120°, 140°, 170°, 280°)$

$\Delta\Theta_i = (0°, 20°, 0°, 30°, 40°), \quad n = 5$

$median(\Delta\Theta_i) = 30°$

Figure 7: Schematic of quantifying alignment

Schematic demonstrating how alignment of cell i is computed. In the first instance (equations shown on the left-hand side below the image) $r_{min} = 0, r_{max} = 1$ and cell i has too few neighbours. In the second instance (equations shown on the right-hand side below the image) $r_{min} = 0, r_{max} = 2$ and cell i has five neighbours (including itself). The orientations of these neighbours are listed in a vector below. Then the nematic angles between cell i and these neighbour orientations are listed below. Finally, the median of these angles is given, which here is $30°$. This means that the alignment of cell i in the neighbourhood [0,2] is $30°$.

For each cell, the median angle of deviation is used as opposed to the mean in order to be less sensitive to outliers. Measurements for each cell are only considered valid if the neighbourhood contains at least five neighbours (including itself) i.e. higher-order alignment in very low confluence, is not meaningful. A schematic demonstrating alignment over different neighbourhoods is given in Figure 8.

The metric demonstrates how the aligning CAFs self-organise over time, and that short-range alignment (0-50μm) is stronger than long-range alignment (150-200μm) (Figure 9). Towards the end of the assay, the alignment process begins to slow down. This could indicate that the system becomes crowded at very high confluence where cells must act in an increasingly cooperative and therefore slower manner (Sadati *et al.*, 2013). Further, the non-aligning CAFs fluctuate moderately in their organisation but never reach alignment. These fluctuations hint at the instability of the non-aligning CAFs and their potential to align under certain conditions.

These experiments raise a number of fundamental questions. Firstly, what are the inherent differences between aligning and non-aligning CAFs that enable them to produce such different emerging behaviour? Secondly, could either alignment or disorder be considered the “default” emergent pattern, and if so, why? Finally, with this information, how easy would it then be to modify cellular properties to change the emergent behaviour? In order to answer these questions, computational modelling was employed, informed by simpler experiments. Starting by considering single cell behaviour, the model could then be built up, moving on to understanding interactions between two cells and finally returning to the complexity of the original assay, involving many cells proliferating to confluence. In this way, the complexity of the system could be broken down into comprehensible components.

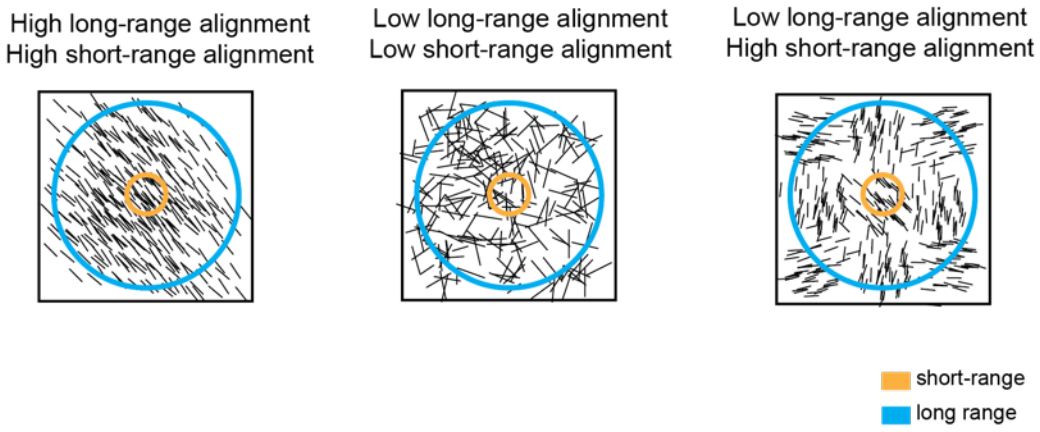


Figure 8: Schematic of alignment

Schematic of example matrix patterns demonstrating different configurations of high and low long-range alignment (LRA, blue) and short-range alignment (SRA, yellow).

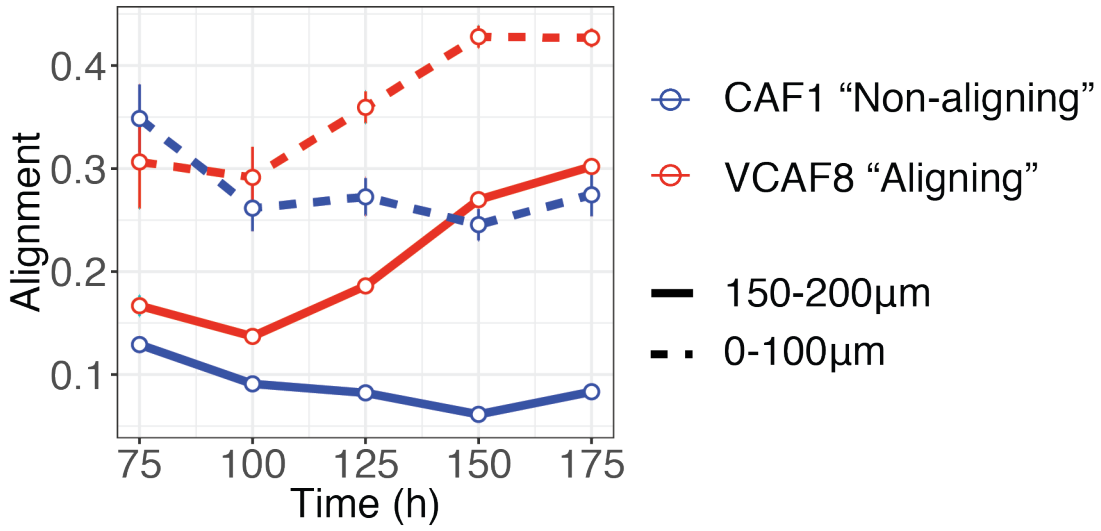


Figure 9: Quantifying alignment of CAFs

Quantification of alignment in time-lapse imaging of aligning (red) and non-aligning (blue) CAFs over 175 hours, at different neighbourhood radii. Aligning CAFs become gradually more aligned from sub-confluence to confluence and alignment is greater over a shorter range (dotted line, $r_{min} = 0\mu m, r_{max} = 100\mu m$) than over a longer range (solid line, $r_{min} = 150\mu m, r_{max} = 200\mu m$).

2.3 Individual cell behaviour

2.3.1 Negative results

Experiments were conducted at low confluence and for short time periods (less than 24 hours) in order to observe cells when proliferation or cell-cell contact events were extremely rare. CAFs were fluorescently labelled with nuclear GFP, imaged and tracked over 24 hours. No statistically clear difference was seen in cell area, aspect ratio or proliferation rate when comparing aligning to non-aligning CAFs (Figure 10, Table 2).

Table 2: Comparing single cell parameters

	Aligning CAF (VCAF8)	Non-aligning CAF (CAF1)
Area	10496μm^2	3310μm^2
Aspect ratio	1:3.5	1:2.7
Speed mean $(\mu m/h)$	54	136
Speed standard deviation $(\mu m/h)$	58	89
Doubling rate (h)	30	20

2.3.2 Migratory persistence

To avoid bias in computing cell persistence due to cell collisions, cells were tracked at low levels of confluence. The program Metamorph was used for cell tracking. *A number of time windows W were chosen over which persistence should be computed, where W is a multiple of Δt.* The persistence over a given window W was taken to be the mean directionality ratio of all cells and is described below: The persistence of a cell $\boldsymbol{i}$ at time $\boldsymbol{t}$ over a window $\boldsymbol{W}$ is defined as

$$P_{i,t}(W) = \frac{\left|\underline{x}_i(t) - \underline{x}_i(t-W)\right|}{\sum_{n=1}^{N}\left|\underline{x}_i(t-(n-1)\Delta t) - \underline{x}_i(t-n\Delta t)\right|} \quad \textbf{(10)}$$

where, $N = \frac{W}{\Delta t}$. Then the persistence of cell i across the entire assay finishing at time t_{max} and is calculated as

$$P_i(W) = \frac{1}{t_{max} - W} \sum_{t=W+\Delta t}^{t_{max}} P_{i,t}(W) \quad \textbf{(11)}$$

Finally, the persistence of all cells through the whole assay is taken as the average across cells:

$$P(W) = \frac{1}{N} \sum_{i=1}^{N} P_i(W) \quad \textbf{(12)}$$

In this work, assays typically ran for 7.5 days (180 hours), from which 720 frames were taken at fifteen minute intervals, therefore $t_{max} = 720$. Windows of 0.5, 1, 2 and 4 hours were employed corresponding in frames to $W = (2,4,8,16)$.

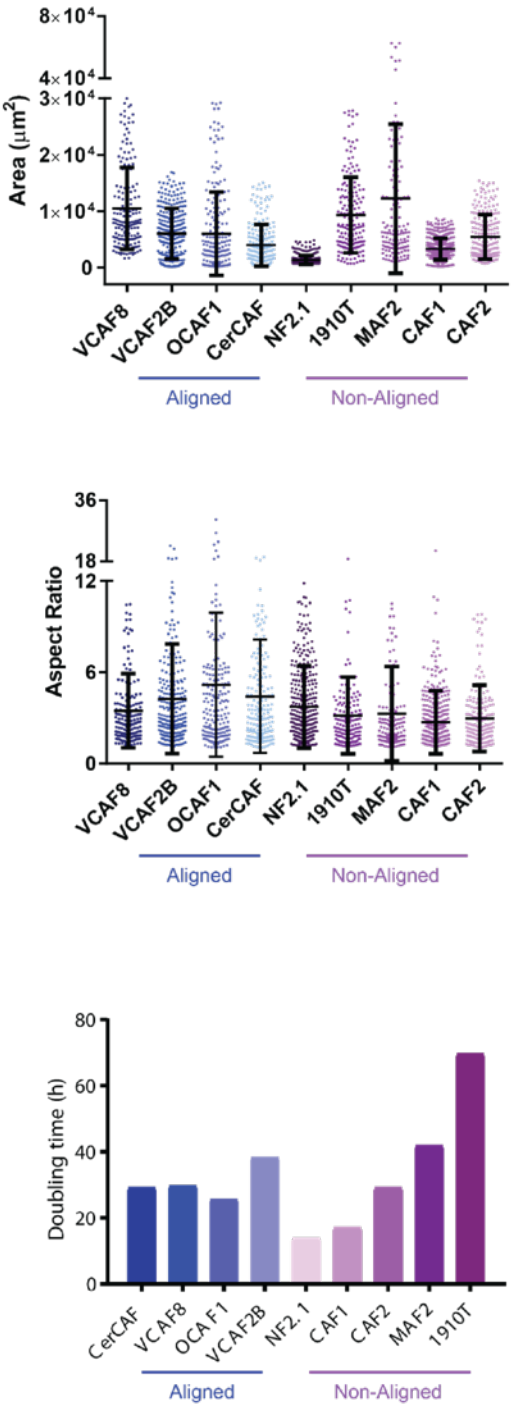


Figure 10: Negative results from comparing aligning and non-aligning CAFs

Analysis of (from top to bottom) cell area, aspect ratio and doubling time in a number of fibroblast lines. The fibroblast lines are classified according to whether they align or not. There is no statistically clear relationship between any of these three cell properties and alignment. This data was collected by Danielle Park.

Aligning CAFs were found to be more persistent than the non-aligning CAFs in sub-confluent conditions over increasing distances (p=1.276e-07, p=4.145e-06, p=0.001126 and p=0.009838 for average distance covered by aligning CAFs in 1, 2, 4 and 8 hours respectively) (Figure 12).

It should be noted that directionality ratio, whilst the most frequently used metric of persistence, can display bias due to differences in cell speed. The non-aligning CAFs move approximately twice as fast as the aligning CAFs. To account for this, measurements were normalized between the two cell types. The positional points of the non-aligning CAF1 cells over a time window were compared with the positions of the cells over a time window of twice as long to account for approximately the same distance covered. The persistence of the aligning cells was then computed on “thinned” positional points of, removing every other point in order to have the same effective distance sampling rate between the two CAFs. This is demonstrated schematically in Figure 11. With this normalisation, aligning CAFs have statistically clearly higher persistence even when normalised for cell speed. In Figure 12, this normalisation is considered. In the course of this work a “biologist-friendly” R script for computing persistence was developed and is now widely used by the Tumour Cell Biology Laboratory for different projects.

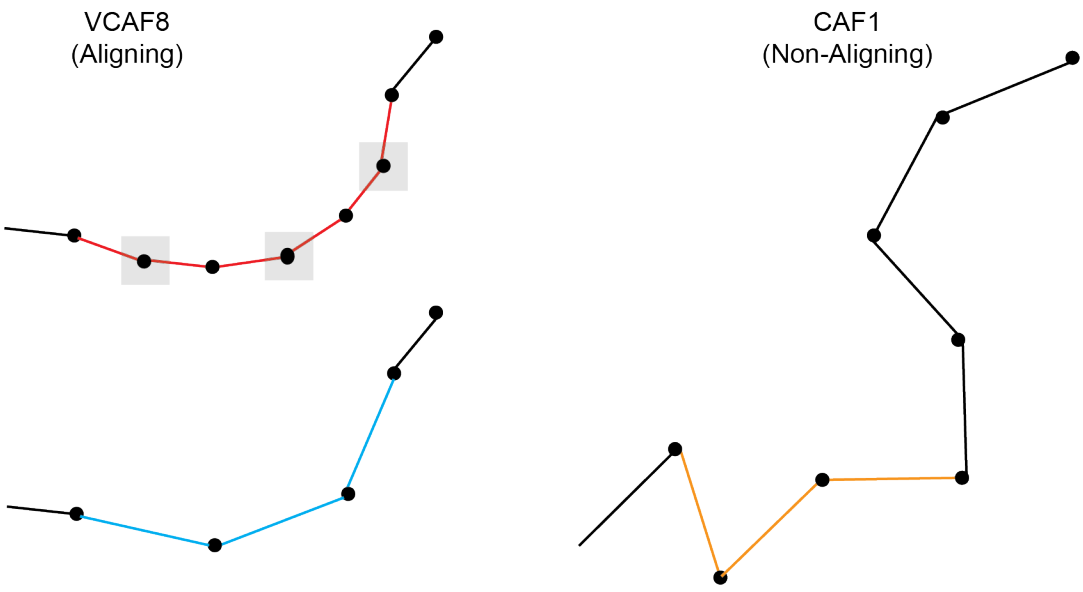


Figure 11: Schematic of persistence normalisation

Non-aligning cells move twice as fast as the aligning cells, so the distance covered by the aligning cells in six time steps (red) is approximately equal to the distance covered by the non-aligning cells in three time steps (orange). The data for the aligning CAFs is thinned by removing every other cell position (highlighted in grey boxes) to create a thinned trajectory (blue) which is compared to the non-aligning trajectory over the same distance (orange).

2.4 Cell-cell interactions

2.4.1 Experiments observing cell-cell collisions

Having studied the key mechanistic differences between aligning and non-aligning CAFs at the level of the single cell, the next step was to consider how differences in cell-cell interactions could alter emergent cell patterns. As shown in **Error! Reference source not found.**, emergence of alignment can be seen at early time points. To explore if collision behaviour could play a role in early emergent behaviour, fibroblast collisions were recorded in sub-confluent conditions. Some of the collisions were recorded by Danielle Park and Samantha George, a summer student in the Tumour Cell Biology laboratory.

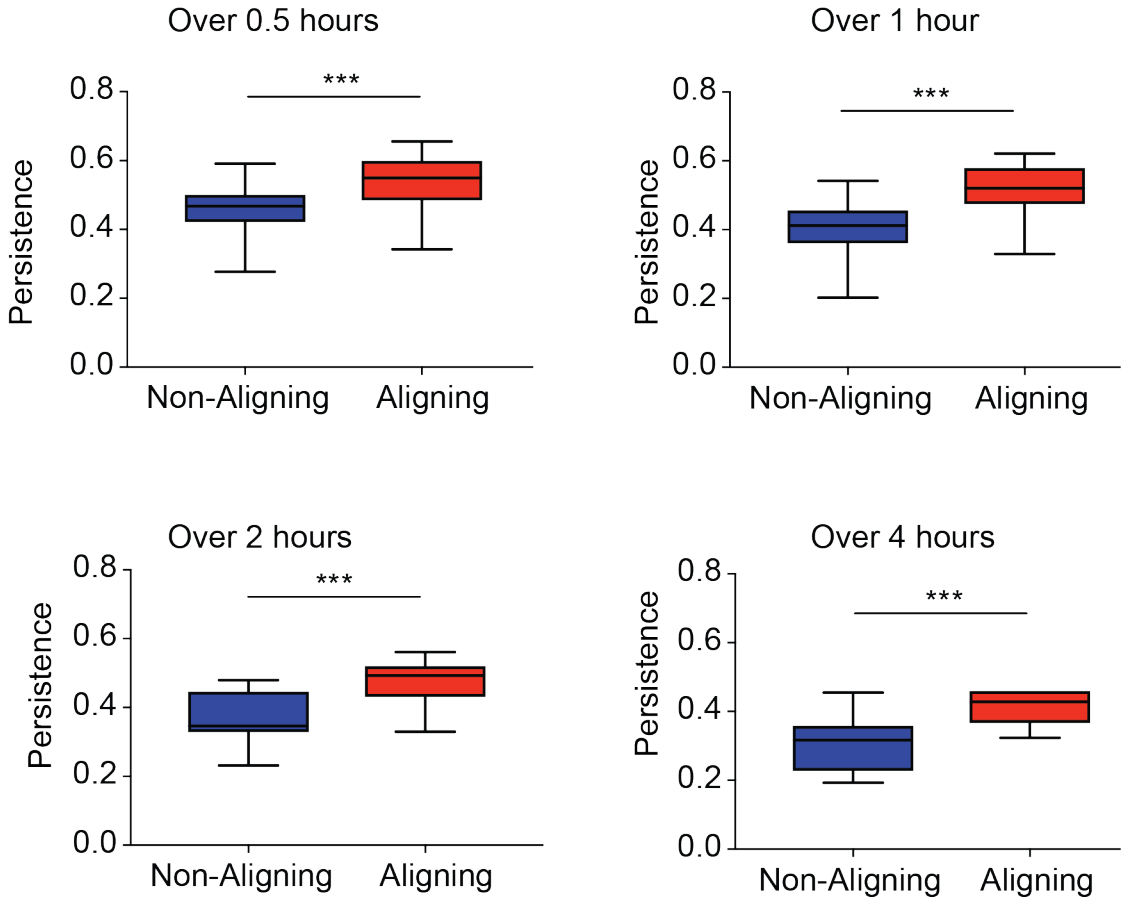


Figure 12: Quantifying persistence of CAFs

Aligning CAFs (red) have enhanced migratory persistence as compared to non-aligning CAFs (blue). Plots show the average persistence of CAFs in sub-confluent conditions over normalized windows for distances travelled by aligning CAFs in 0.5, 1, 2 and 4 hours.

2.4.2 Rules of collision tracking and analysis

For a collision to be valid, no more than two cells could come into contact for the duration of the collision, and neither could undergo mitosis.

Cell orientation was manually recorded using ImageJ. Unfortunately in these experiments, cells were not nuclear labelled, rendering automatic tracking difficult. Instead, cell orientation was manually recorded. The frame rate of these experiments was five minutes. In the three frames (fifteen minutes) before the collision, the directional migration of the cell was observed. In this short time interval even the non-aligning cells moved in an approximately straight line (ie the persistence length was generally greater than fifteen minutes). A line was drawn from the tail of the cell to the head of the cell in the direction of the cell's migration. In this way, the orientation of the cell before the collision was recorded. Similarly the orientation of the cell after collision was recorded by manually observing the cell for three frames after the cell-cell contact was broken. Very often in collisions, one cell is more motile over the time-scale of the collision. In the case where both cells were active in the collision, two collisions were recorded with each cell being the cell of interest respectively.

The method used for measuring cell orientation before and after collision could be improved in future work. There is evidence that the orientation of cell protrusions correlates poorly with the cell's direction of migration (Yolland *et al.*, 2019). It would be beneficial to run similar experiments with nuclear labelling, to establish the cell's migrational orientation automatically as described in 2.2.1.

Considering first just the repolarisation of the cells of interest after collision shows a statistically significant decrease in repolarisation of the aligning CAFs as compared to the non-aligning CAFs (Figure 13). Analysis shows suppressed CIL response in aligning fibroblasts (n=90 collisions in total from three independent experiments. $p=5x10^{-5}$, one-sided t-test).

Whilst this result indicates a difference in response of cells upon collision, inherent to alignment is the requirement for two cells to coordinate their behaviour. A dampened repolarisation response as shown in the aligning CAFs (Figure 14A) is not necessarily indicative of alignment. The system only moves closer to alignment if the two cells have more similar orientations to each other after the collision has taken place (see Figure 14A). To reflect this, a more in-depth analysis was carried out considering the angle of a cell before and after a collision relative to the orientation of the cell it collided into.

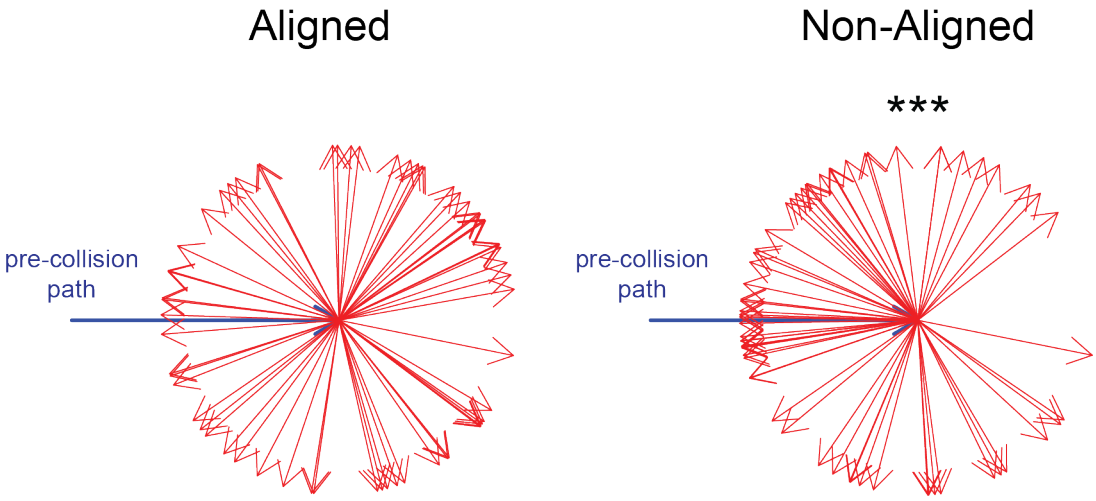


Figure 13: Roseplots showing CAF CIL response

Analysis of CIL response in aligned (VCAF8, VCAF2B) and non-aligned (CAF1, NF2.1, CAF2) fibroblasts. Change in cell trajectory upon collision relative to the pre-collision path (blue arrow). These collisions were recorded by the author, Danielle Park and Samantha George.

Collision guidance events were defined as collisions in which the cell of interest re-orientated to be closer to the cell it collided with; if the angle of approach minus the angle of departure (relative to the reference cell) gave a value in the range of 10-90 degrees. Figure 14B encapsulate this analysis: with the angle of approach and departure relative to the neighbouring cell plotted on the x and y axes, respectively and the size of the circle indicating extent of cell repolarisation. A large number of collisions of the aligning CAFs fall into the region of collision guidance, where the angle of departure is less than the angle of approach and the repolarisation is less than 90°. This is visible in Figure 14B as small circles below the line representing equivalent approach and departure angles (inside blue triangle). The proportion of collisions falling within the collision guidance category is statistically clearly higher for the aligning CAFs than the non-aligning CAFs (p-value = 0.006, z-test, see Appendix).

Traditionally, CIL has been measured solely in terms of a single cell (Desai *et al.*, 2013; Davis *et al.*, 2015), however it is clear that this this does not elucidate the role of cell-cell interactions in the context of higher-order behaviour. The approach to collision analysis developed here, considering CIL of a cell with respect to its

neighbour demonstrates how collision guidance events underpin the emergence of anisotropy.

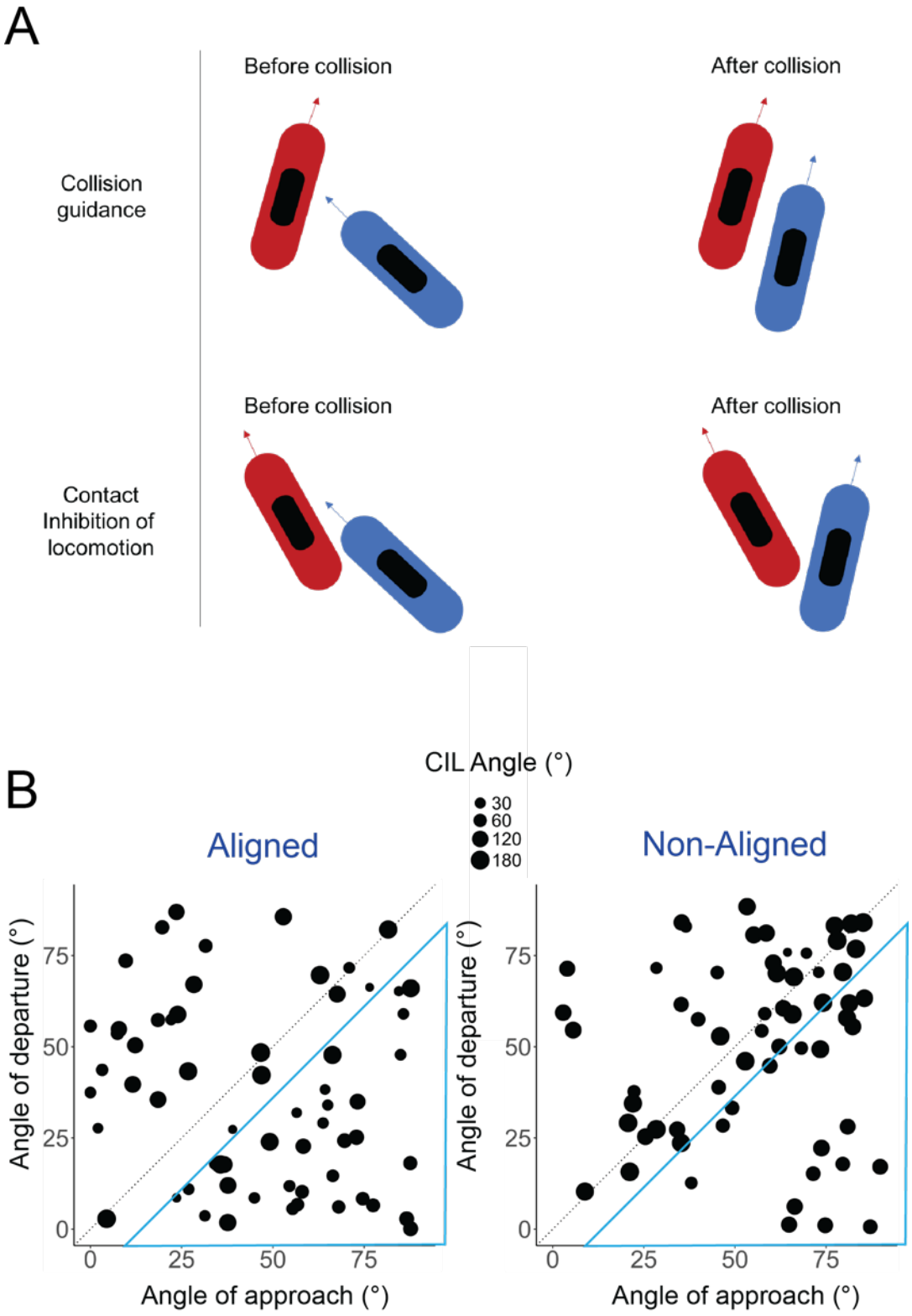


Figure 14: Measuring cell-cell collision guidance

Two cells need to coordinate their behaviour for alignment to occur. (A) In both the top and bottom panels, the blue cell undergoes the same change in orientation, but can lead to improved alignment (collision guidance) as shown in the top panel, or a worsening of alignment (contact inhibition of locomotion) as shown in the bottom panel. (B) Analysis of collisions, plotting the angle of approach (x axis) and departure (y axis) relative to the cell it collides with. Collisions in which angle of departure - angle of approach is in the range $[\mathbf{10}^\circ, \mathbf{90}^\circ]$ (inside blue triangle) reflect collision guidance events. Size of circle indicates the cell's change in orientation as shown in Figure 13. Aligning CAFs have significantly more collision guidance events (p=0.006, z-test, n=124 collisions).

Importantly, more detailed analysis of these collisions revealed that in the majority of collisions, the faster moving cell (called the "Altering cell") changed direction by a greater amount than the slower moving cell (called the "Dominant cell") (Figure 15, n=20 pairs of cells, p<0.0001, two-tailed t-test). This is the opposite of what would be expected in a collision between two inert objects of approximately equal mass and suggests fibroblasts are not undergoing inelastic collisions, but rather undergo an orchestrated response to biological signalling informing the collision behaviour.

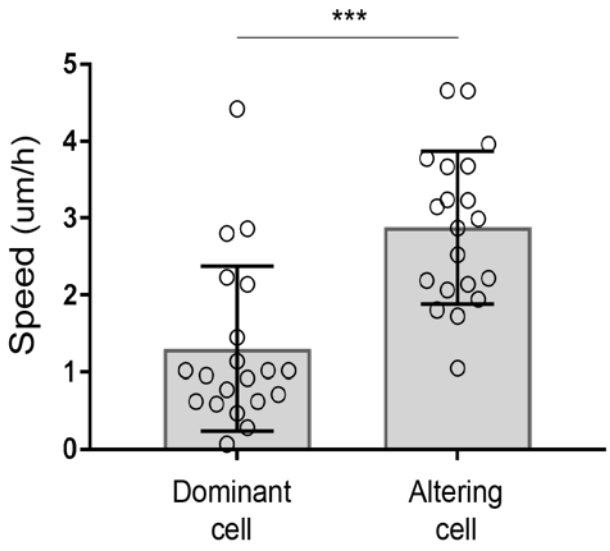


Figure 15: Speed of cells before collision

Cell speed prior to collision in both the altering cells (cell undergoing collision guidance) and dominant cells (n=20 pairs of cells, p<0.0001 two-tailed t-test, bars indicate mean and SD).

2.5 Model construction

To explore how CAFs are able to coordinate their behaviour to generate alignment over several orders of magnitude, a mathematical model was established, adapted from the Vicsek flocking model (Vicsek *et al.*, 1995; Chaté *et al.*, 2008). The

change in the orientation of a cell is computed as a weighted function of two terms: individual migratory noise and cell-cell collision guidance. A fibroblast's individual migratory noise reflects its persistent migration as a result of cell polarisation (Gorelik and Gautreau, 2014) and the cell-cell guidance mechanism determines the extent of "flocking" behaviour.

2.5.1 Model overview

It is perhaps useful to begin by considering a schematic version of the model, which whilst not mathematically rigorous, aims to help the reader to conceptualise the different components of the flocking model. In its schematic form, the model describing the orientation of a cell i at time $t + \Delta t$ can be written as

$$\theta_i(t + \Delta t) = w_p[\theta_i(t) + \eta_i(\Delta t)] + w_c f\left(\theta^i_{cells}\right), \qquad \textbf{(13)}$$

$$w_p = 1 - w_c, \qquad 0 \leq w_c \leq 1.$$

The first term indicates the individual migration of the cell in the absence of external influences. The $\eta_i(t)$ denotes the noise of cell i at time t in its persistent migration. The function $f\left(\theta^i_{cells}\right)$ is a function describing how cell i will flock with its neighbouring cells. The weight w_c describes how much the orientation of the cell will be influence by cell-cell flocking. Since $w_p = 1 - w_c$, the two free parameters in the model are η_i and w_c. The smaller the value of w_c, the larger the value of w_p (the weight controlling persistence), meaning a cell's orientation will be determinined to a greater extent by its individual migratory persistence than flocking with its cell contacts.

Unlike the original Vicsek model, cells are able to coordinate their behaviour nematically (Ginelli *et al.*, 2010; Marchetti *et al.*, 2013) and the extent of coordination could be explicitly controlled by the cell-cell guidance term. The rest of the model description describes this model more completely, expanding the definition of the flocking function f and reducing the orientation of a cell into its X and Y components, which act as update formulae to account for issue that arise with resolving the trigonometric quadrant.

2.5.2 The model

At time zero, cells were placed at random and each cell assigned a random orientation. Fibroblasts are modelled as a node representing the head of the cell, followed by two nodes of twice the diameter representing the cell body, followed by another smaller node for the tail. This diamond-like shape has aspect ratio 1:3 and reflects one of the typical cell morphologies. For a cell i at time t, $\theta_i(t)$ denotes its orientation, s_i denotes its speed and $\left(x_i(t), y_i(t)\right)$ denotes its position.

Cell speed

At initialisation, a cell i is given a constant speed $\mathrm{s_i}$ drawn from a Gaussian distribution $\mathrm{S}\sim\mathrm{N}(32\mu\mathrm{m}, 4\mu\mathrm{m})$ fitted to the experimental data. Each cell was given a fixed speed, similar to the Vicsek model but unlike the Vicsek model, the speeds between cells could be different, reflecting the normal distribution of cell speeds observed in the data. Additionally, whilst the mean and standard deviation of the speed distribution meant that negative values were extremely unlikely to be drawn, in such an event, the absolute value of $\mathrm{s_i}$ was taken so that cells would always be assigned a non-negative speed.

Individual migratory noise

Cells move along their long axis at a constant speed and the individual migratory noise of a cell is modelled as a persistent random walk $\theta_\mathrm{i}(\mathrm{t}) + \eta_\mathrm{i}(\Delta\mathrm{t})$, where $\eta_\mathrm{i}(\Delta\mathrm{t})$ is Gaussian distributed with mean zero. Specifically, noise is dependent on time step $\Delta\mathrm{t}$ such that

$$\eta_\mathrm{i}(\Delta\mathrm{t})\sim\mathrm{N}\left(0, \eta\sqrt{\Delta\mathrm{t}}\right)$$

Here $\Delta\mathrm{t}$ is used to describe the frame rate and in this work the default frame rate for experiments is fifteen minutes. The positional X and Y-components of this term for cell i at time t can be defined as

$$\begin{aligned} \mathrm{X_{i,p}(t)} &= \cos\left(\theta_\mathrm{i}(\mathrm{t}) + \eta_\mathrm{i}(\Delta\mathrm{t})\right), \\ \mathrm{Y_{i,p}(t)} &= \sin\left(\theta_\mathrm{i}(\mathrm{t}) + \eta_\mathrm{i}(\Delta\mathrm{t})\right). \end{aligned} \quad \textbf{(14)}$$

Cell-cell collision guidance

If a cell i is in direct contact with N cells, $N > 0$, then the X and Y-components of the effect of cell guidance on i at time t are defined as

$$\mathrm{X_{i,c}(t)} = \frac{1}{\mathrm{N}}\sum_{\mathrm{j=1}}^{\mathrm{N}} \cos\left(\tilde{\theta}_{\mathrm{j}}(\mathrm{t})\right), \qquad \mathrm{Y_{i,c}(t)} = \frac{1}{\mathrm{N}}\sum_{\mathrm{j=1}}^{\mathrm{N}} \sin\left(\tilde{\theta}_{\mathrm{j}}(\mathrm{t})\right), \tag{15}$$

where

$$\tilde{\theta}_{\mathrm{j}}(\mathrm{t}) = \begin{cases} \theta_j(t) & \text{if } |\, \theta_j(t) - \theta_i(t) \,| \ (\mathrm{mod}\ \pi) < \frac{\pi}{2}, \\ \theta_j(t) + \frac{\pi}{2} & \text{otherwise.} \end{cases} \tag{16}$$

This adaptation constitutes the ability of fibroblasts to align in a nematic manner. For greater tractability, an assumption of the model is that w_c remains fixed for any non-zero value of N, so that, provided a cell is in contact with at least one other cell, the degree of flocking will be fixed. This means cell-cell guidance is independent of N for $N > 0$. The angle of cell-cell flocking is an average of all N of a cell's contacts. However, the confluence levels both *in silico* and *in vivo* mean that N is kept small in practice. Example outcomes of collisions depending on the parameter values of individual migratory noise and cell-cell collision guidance are shown in Figure 16.

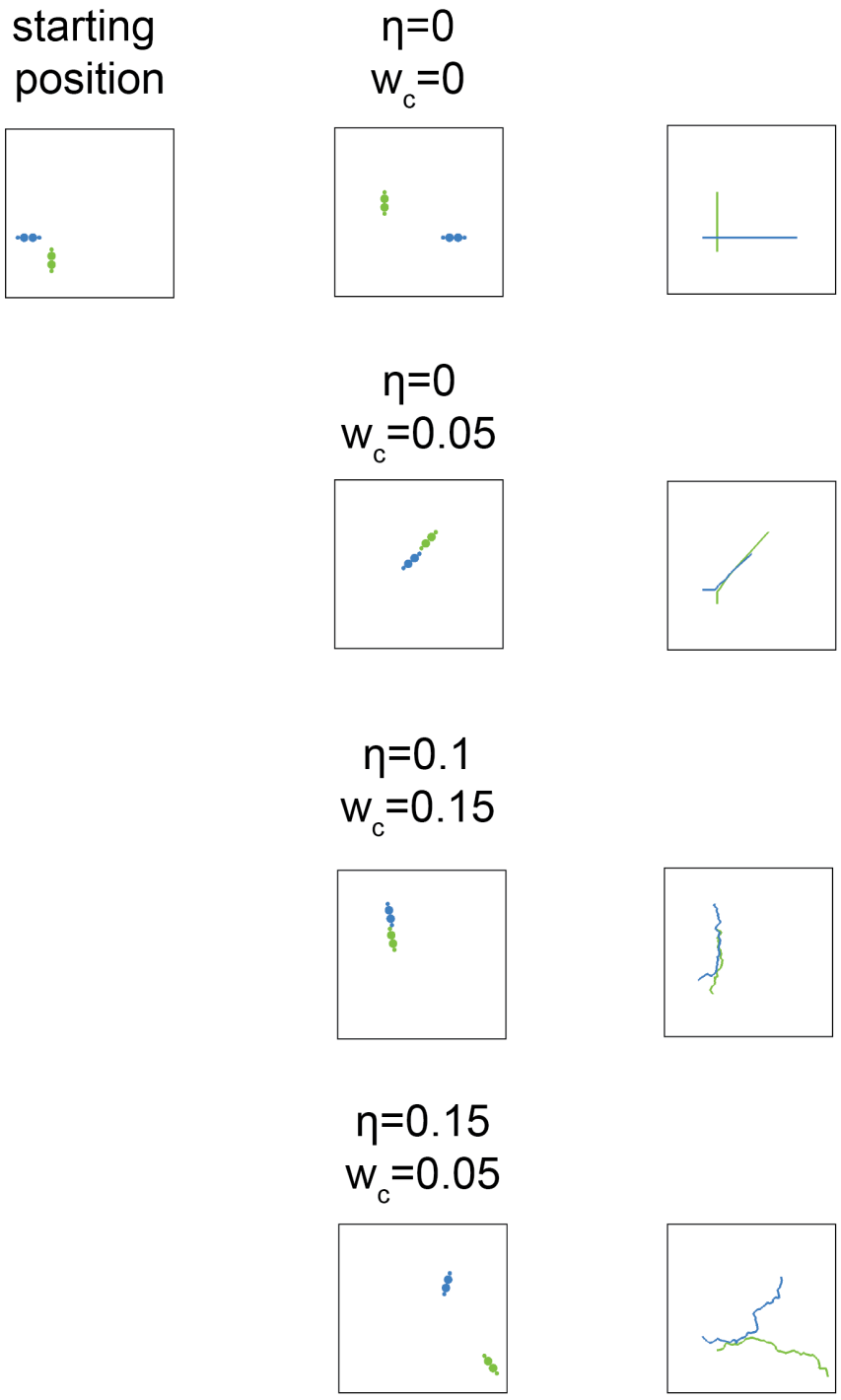


Figure 16: Example collision outcomes

The panel on the left shows the positions of two cells before they collide. The middle column shows the positions of the same cells after a collision, for different indicate values of individual migratory noise (η) and cell-cell collision guidance (w_c). The right-most column shows the cell trajectories for the duration of the collision.

Computing cell orientation

Having computed the individual migratory noise and cell-cell collision guidance, the X and Y-components of cell i at time t are then written as a weighted function of the two terms:

$$X_i(t) = \frac{1}{w_p + w_c}\left(w_p X_{i,p}(t) + w_c X_{i,c}(t)\right), \quad \textbf{(17)}$$
$$Y_i(t) = \frac{1}{w_p + w_c}\left(w_p Y_{i,p}(t) + w_c Y_{i,c}(t)\right),$$

where $w_p = 1 - w_c$, $0 \leq w_p, w_c \leq 1$. If a cell is not in contact with any other cell, then w_c is set to zero. Typically, in Vicsek-like models (Vicsek *et al.*, 1995; Ginelli *et al.*, 2010), averaging between neighbours is instantaneous. However, this is not reflective of true cell behaviour, where alignment is more gradual. The advantage of having a weighted function is that the collision guidance mechanism can be varied to reflect this.

The orientation of cell i at time $t + \Delta t$ is then computed as

$$\theta_i(t + \Delta t) = \tan^{-1}\left(\frac{Y_i}{X_i}\right) \quad \textbf{(18)}$$

which must then be adjusted for quadrant of the arctan function so that

$$\theta_i(t + \Delta t) = \begin{cases} \theta_i(t + \Delta t) \text{ if } X_i \geq 0, Y_i \geq 0, \\ \theta_i(t + \Delta t) + \pi \text{ if } X_i < 0, \\ \theta_i(t + \Delta t) + 2\pi \text{ if } X_i \geq 0, Y_i < 0. \end{cases} \quad \textbf{(19)}$$

Finally, cell position is updated so that

$$\begin{bmatrix} x_i(t + \Delta t) \\ y_i(t + \Delta t) \end{bmatrix} = \begin{bmatrix} x_i(t) + s_i \Delta t \cos\left(\theta_i(t + \Delta t)\right) v_e \\ y_i(t) + s_i \Delta t \sin\left(\theta_i(t + \Delta t)\right) v_e \end{bmatrix} \quad \textbf{(20)}$$

where v_e is a proxy for volume exclusion defined by

$$v_e = \begin{cases} 0.25 \text{ if the head node of a cell is overlapping with any other node of another cell,} \\ 1 \qquad \text{otherwise} \end{cases}$$

The distance between cells at which this pseudo-volume exclusion is activated is set by the user (Figure 17). This is to allow for a small amount of overlap between cells, which is observed as cells deform their shapes to slide past another cell. The "hard" inner region can be thought of the cell's nucleus which is less deformable. The two free parameters of this volume exclusion mechanism are the size of the hard inner region and then value of v_e when volume exclusion is triggered. Unless otherwise indicated, in all simulations, the inner region was defined as 75% of a cell's area and the value of v_e was set to 0.25. A high value for $\mathrm{v_e}$ could lead to the system becoming jammed (Sadati et al., 2013), however for the chosen values in this work, the mechanism serves only to reduce cells overlapping or going through each other and has parameter values that do not lead to jammed behaviour. A schematic of the model, showing the mechanisms of individual migratory persistence and cell-cell collision guidance is given in Figure 18.

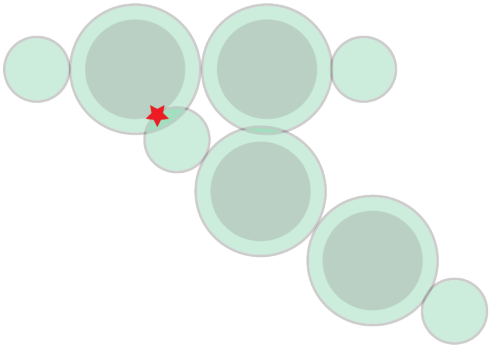


Figure 17: Implementing volume exclusion term

Schematic showing the pseudo-volume exclusion mechanism in the model. A cell's volume exclusion is activated if its head overlaps with an inner region (indicated in grey) of another cell (indicated with a red star). The size of the inner region is defined by the user but in all simulations presented in this work is set to 75% of the size of the body nodes (indicated). When volume exclusion is activated the cell will travel 25% of the normal distance covered in a single time step.

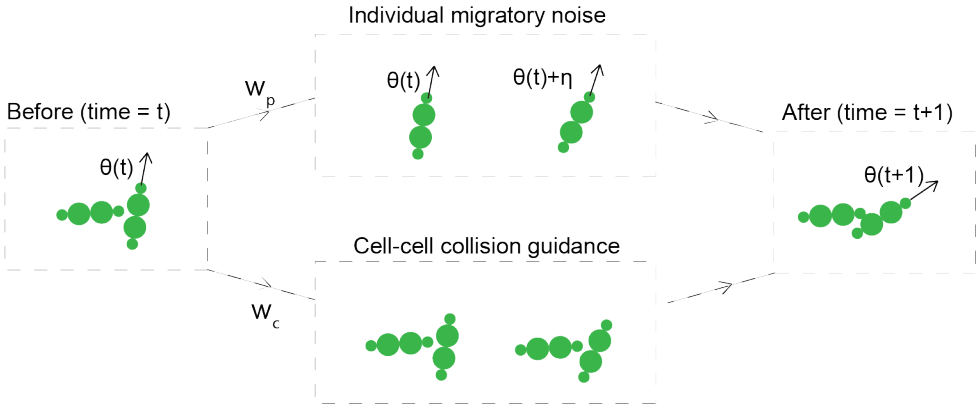


Figure 18: Model schematic I

Model schematic showing how the orientation of a cell changes as a function of individual migratory noise and cell-cell collision guidance. Here only the change in orientation of the cell on the right of the collision is shown for clarity.

2.5.3 Additional model features

Cell position

The positional coordinates of a cell i at time t are denoted $(x_i(t), y_i(t))$ and represent the centre of the cell's head node. When a cell's orientation changes, the cell turns about its centre of mass, which is determined based on the cell's aspect ratio.

Periodic boundary conditions

Periodic boundary conditions are used throughout the simulations. Topologically, this can be thought of as mapping the two-dimensional simulation space onto a torus (Grossman, Aranson and Ben Jacob, 2008). This meant that the Euclidean distance between two cells i and j in a simulation with dimensions of length l (specified by the user) would be defined as

$$d(i,j) = \sqrt{\left(\min\left(|x_i - x_j|, l - |x_i - x_j|\right)\right)^2 + \left(\min\left(|y_i - y_j|, l - |y_i - y_j|\right)\right)^2}. \quad \textbf{(21)}$$

Voxel grid

A non-physical voxel grid was implemented in the model to implement cell proliferation and increase computation speed (described below). The simulation

space was divided up into grid squares of equal size. The number of grid squares was defined as:

$$\left\lfloor \frac{\text{simulationWindowSize}}{2 * \text{cellAspectRatio} * \text{radiusCellNodes}} \right\rfloor, \quad \textbf{(22)}$$

meaning that the size of each grid square was just larger than the length of a cell. At each time step, the voxel grid square in which the centre of each cell's head node fell into was recorded.

Proliferation

Cell proliferation is implemented using the Voxel grid. Cell density at time t can be defined to as:

$$\rho(t) = \rho(0) \times 2^{\frac{t}{d}} \quad \textbf{(23)}$$

where $\rho(0)$ denotes the initial cell density at time 0 and d is the average time it takes for cell population density to double. At each time step, if the number of cells in the simulation is less than $\rho(t)$, then a cell i is selected at random to be the "mother" cell. If one or more of the neighbouring voxel grid squares to this cell are empty, a new daughter cell j will be initialised in a one of these empty grid squares, selected at random. The orientation of the new daughter cell will be opposite to that of its mother so that

$$\theta_j(t) = \theta_i(t) + \pi. \quad \textbf{(24)}$$

If there is no empty neighbouring voxel grid square, a new candidate mother cell is chosen and so on until there is no space remaining within the entire simulation area. This method simulates the phenomenon known as contact inhibition of proliferation (Puliafito *et al.*, 2011).

In the vast majority of the simulations, cell density was kept constant, and so we did not investigate the effects of the proliferation mechanism in detail. Other models have considered this in more detail, with cell's dividing according to an internal clock or with more detailed dynamics of division and it would be interesting to investigate this further in the future. However, simulations were run with very high

proliferation rates to see if this could jam the system and it was found that typically aligning cells would still align, suggesting that proliferation is not a primary parameter in the model. The doubling rate d was fitted to the experimental data (now indicated in the text), so that over numerous time points, it was known what the cell density ought to be. For this reason, if a mother cell was not able to divide, a new candidate mother was chosen at random instead of aborting. The initial orientation of the daughter cell was set to be the opposite direction to the orientation of the mother cell. This was done to simulate spindle division and the observed movement of daughter cells away from each other in experiments.

A complete list of free parameters of the model is given in
Table 3. Unless otherwise simulations take parameters with the default values given in the "Typical values" column. The primary parameters that will be varied are individual migratory noise (η) and cell-cell collision guidance (w_c).

Table 3: List of parameters I

Parameter	Meaning	Typical value
ρ	Number of cells in simulation	800 (corresponding to ~40% confluence)
d	Number of time steps required for population to double	∞ (ie there is no proliferation)
μ	Mean of Gaussian distribution for cell speed	$32\mu m/h$
σ	Standard deviation of Gaussian distribution for cell speed	$4\mu m/h$
a	Number of body nodes (excluding the head and tail nodes), determining aspect ratio	2
r	Radius of cell body nodes (cell area is then $ar^2 + \frac{r^2}{2}$)	$13.5\mu m$ (cell area is then $1410\mu m^2$
η	Individual migratory noise. At each time step $\eta_i(t) \sim N(0, \eta)$	varied

w_c	Weighting of flocking mechanism (collision guidance)	varied

2.5.4 Code structure

Having described the features of the model, it is worth noting the structure of the code used to implement the model (Figure 19). The development and streamlining of the code constituted a large component of this work and provides a comprehensible and tractable framework which can be used by other researchers in the future.

The code is written in C++, known for its object-oriented programming features. The ability to make hierarchical object classes in C++ makes it ideally suited for modelling the tumour microenvironment (TME), where there are multiple cell types (fibroblasts, cancer cells, T-cells etc.), elements that make up individual cells (here called “nodes”, representing the cell head, tail etc.) and the ECM (discussed in the next chapter).

Code was built and verified incrementally through unit testing. Simple simulations were run with one and two cells with designated positions and orientations to check cell behaviour was in line with expectation of the model.

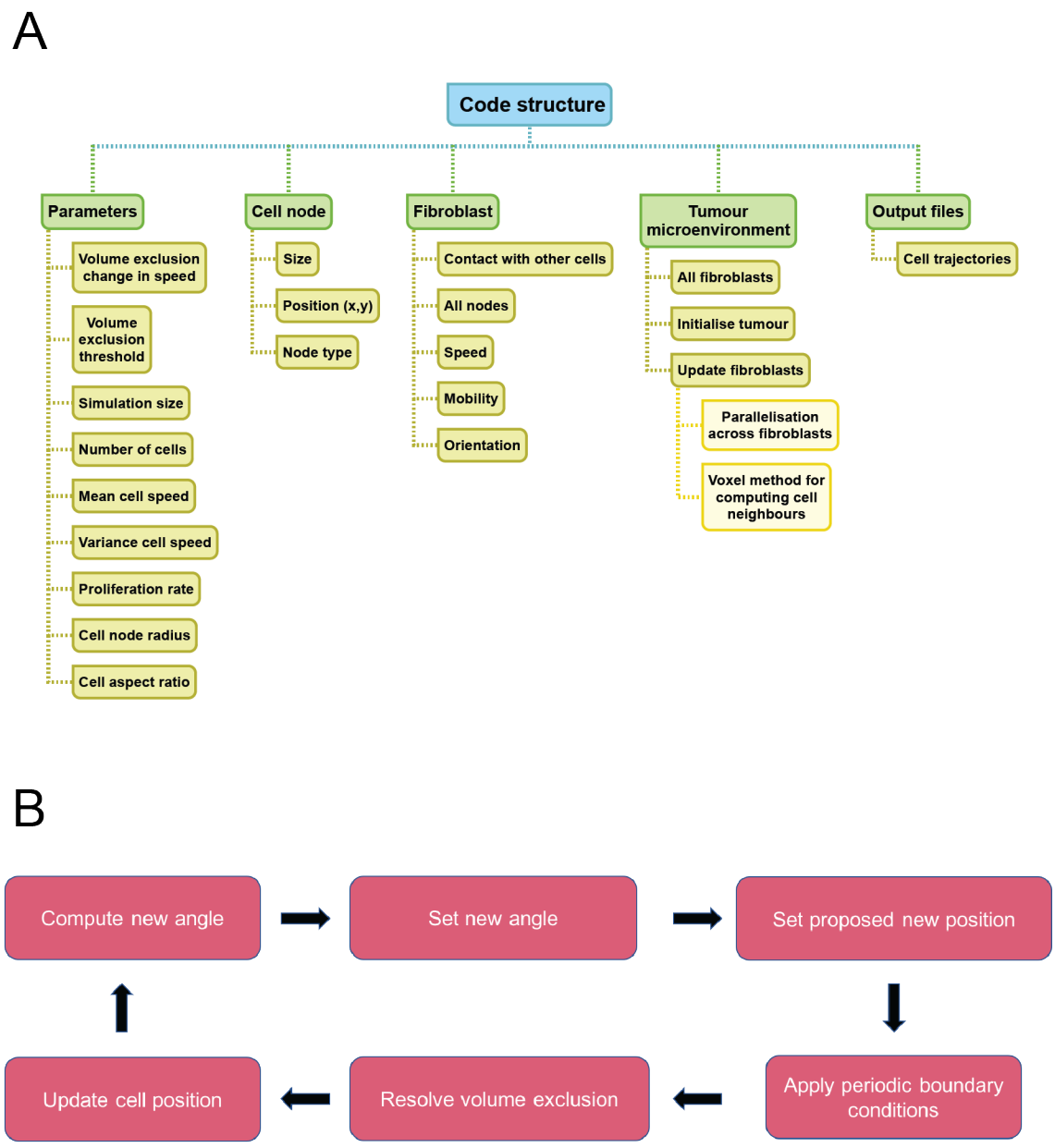


Figure 19: Code and model structure I

(A) Map showing code structure as written in object-oriented programming language C++. Major classes are shown in green, with members of each class shown in dark yellow. Light yellow boxes indicate coding techniques to improve code efficiency. (B) Flow diagram showing the order of events in the update function. Once the new angle of the cell has been determined and set, the proposed new position of the cell is evaluated, first considering the effects of periodic boundary conditions and secondly if the cell's head overlaps with the hard inner area of another cell, in which case the pseudo-volume exclusion term v_e is set to a value of 0.25, and the distance moved by the cell in that time step will be 0.25 of the distance otherwise. This is what is meant by resolving volume exclusion. Once these factors have been considered the cell positions are updated.

As an input, the user provides the model parameters including simulation window size. The simulation is then initialised with all the required cells. Time is modelled discretely such that each time step represents one frame in the experiments initialised and periodic boundary conditions are employed. Properties of each class are indicated in Figure 19.

At each time step, a function “Update fibroblasts” is called. The new angle of each fibroblast is computed, volume exclusion criteria are next resolved and then fibroblast position is updated. This is shown in detail in a flow diagram (Figure 19B). Two coding techniques were implemented to significantly speed up simulation time. Firstly, the “Update fibroblasts” function was parallelised using the application programming interface (API) OpenMP. This allowed individual cell orientations to be computed simultaneously. Secondly, the voxel grid (2.5.3) was defined and cells assigned to a particular grid square at each time point. Instead of computing a distance map between all cells, which can be computationally expensive, the distance between two cells was only computed if they belonged to neighbouring voxel grid squares.

The simulations generate an output file giving cell trajectories. For each time step, cell ID, position and orientation are given. A further C++ script was written, employing OpenGL to then visualise these cell trajectories as videos or individual frames.

2.5.5 Simulation setup

The simulation window size is $1720 \times 1720 \mu m$ and periodic boundary conditions are employed. Time is modelled discretely such that each time step (Δt) represents one frame in the experiments, which corresponds to 15 minutes and the model typically simulates 7-day assays with 800 cells, corresponding to 40% confluence. At time $t = 0$, cells are seeded randomly with random orientation.

As described in section 1.7.7 with the Vicsek model, the choice of time step or cell speed can have an impact on the emergent behaviour. If cell speed is higher or

similarly if the time step is bigger, then a cell will move further between samplings (Dallon, Sherratt and Maini, 1999), resulting in less alignment caused by fewer cell interactions. Likewise, with a smaller time step or slower cells, the cell will interact with the same neighbours many times. However, for relevant and reasonable choices of time step and speed, these factors have only minor effects on the system and do not alter phase transitions (Dallon, Sherratt and Maini, 1999; Baglietto, Albano and Candia, 2012).

CAFs are represented as two large circles with two smaller circles of half the radius at either end. For most simulations, the total cell area is $1410\mu m^2$, corresponding to a large circle radius of $13.5\mu m$ and the length of the cell is therefore $80\mu m$.

2.6 Parameter fitting

2.6.1 Fitting noise

Cell migration is modelled through the allocation of a cell speed and orientation at any given time step [1, 2]. A cell i's orientation, $\theta_{\mathrm{i}}(\mathrm{t})$, at time t, is determined by: $\theta_{\mathrm{i}}(\mathrm{t}) = \theta_{\mathrm{i}}(\mathrm{t}-\Delta\mathrm{t}) + \eta_{\mathrm{i}}(\Delta\mathrm{t})$, with randomness introduced via the Gaussian random variable $\eta_{\mathrm{i}}(\Delta\mathrm{t}) \sim \mathrm{N}\left(0, \eta\sqrt{\Delta\mathrm{t}}\right)$ as described in section 2.5.2, where a single time step is taken to be a unit of fifteen minutes.To quantify η from persistence data, the following procedure was developed:

Step 1: Compute persistence of smoothed cell trajectories

Cells tracked for at least one hour were recorded and a spline tracing the cell's trajectory was produced using the loess package in R $(\alpha = 0.5)$. This was in order to smooth the intra-cellular movement, which resulted in many small fluctuations in the trajectories (Figure 20). For each cell, the median persistence was computed as described in section 2.3.2 for an interval of one hour.

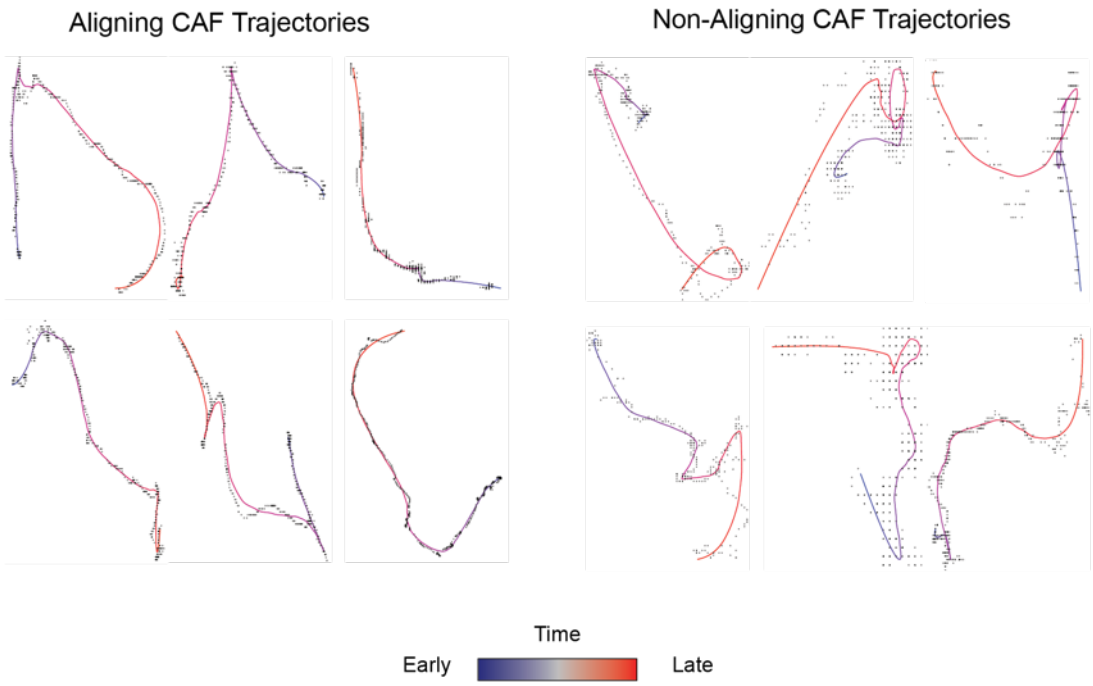


Figure 20: Fitting noise parameter to experimental data

Smoothing cell trajectories. Exact cell locations are shown by dots. A smoothing curve is laid over the trajectories showing a cell's smoothed trajectory through time.

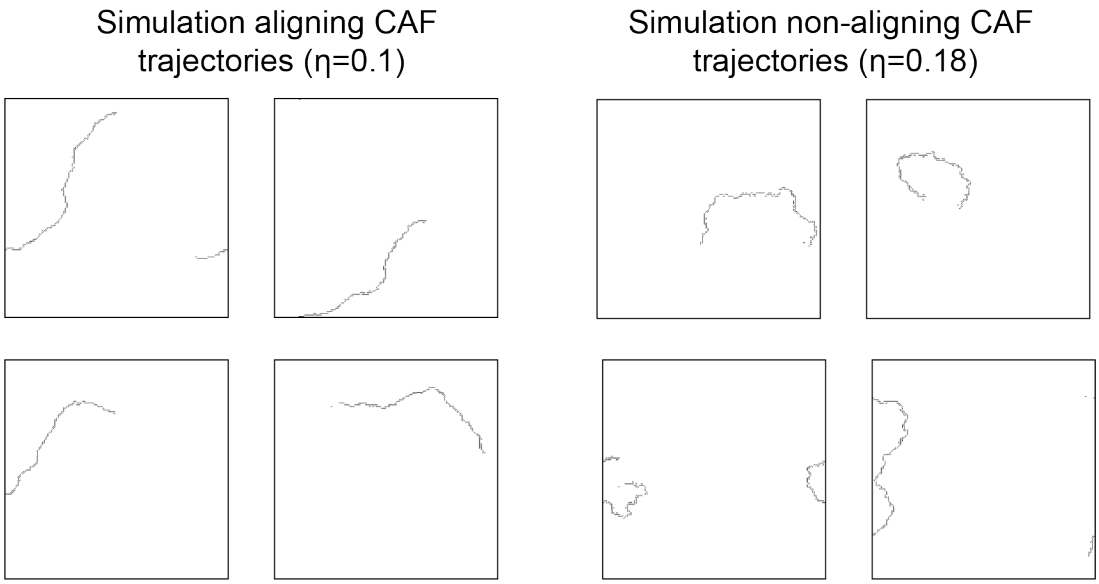


Figure 21: Simulation trajectories with fitted noise

Example cell trajectories with fitted levels of noise for aligning CAFs (left, $\eta = 0.1$) and non-aligning CAFs (right, $\eta = 0.18$). Simulations have periodic boundary conditions, as illustrated in some of the examples.

Step 2: Compute cell speed

In the model, each cell has a fixed speed, determined at its initialisation, which is drawn from a Gaussian distribution. Interestingly, cell tracking revealed that the distance travelled by a cell in one individual time step tended to follow an exponential distribution, often not moving at all in a single time step. In hindsight, the assumption that cells move at constant speed may overlook a key feature of cell motility. In future work it may be of interest to refine the cell speed parameter to account for heterogeneity in single cells using a Bayesian method of inference (Metzner *et al.*, 2015). However, taking the mean speed of each cell across the whole time-course revealed that the distribution of average speed for each cell closely resembled a normal distribution. In simulating each experiment, it was these normal distributions that were used.

Step 3: Run simulations with this cell speed and varying noise

Simulations with fifty cells of radius 0 were run with incrementally increasing levels of noise. This means the cells would completely ignore each other and is equivalent to running fifty independent single cell simulations. Crucially, the cells had the same speed as the experiments. The persistence of these simulated cells over a window of one hour was then computed.

Step 4: Match up simulations and experiments to determine likely level of noise

The experimental results were then matched with simulations to select the most likely value of $Var(\eta)$ given a cell's persistence. The most likely values of noise for the aligning and non-aligning cells were $Var(\eta) = 0.1,$ and $Var(\eta) = 0.18$ respectively. Example *in silico* trajectories of cells are given in Figure 21.

2.6.2 Fitting level of collision guidance

Simulations were run with noise $\eta \sim N(0, 0.1)$, describing the persistence of aligning CAFs for incremental values of collision guidance (w_c) between 0 and 1. The short-range and long-range alignment of these simulations over time was computed and through a least squares approach, compared with the alignment of the experiments to choose the most likely value of collision guidance that would match the simulations to the experiments. This value of collision guidance is $w_c = 0.04$.

2.6.3 Challenges of fitting parameters and making model predictions

There were several challenges in fitting the parameters of the model to the experimental assays. Firstly, in assuming that cells move according to a persistent random walk, the model neglects to consider random cell repolarisations (Metzner *et al.*, 2015). In future work, it would be of interest to quantify the frequency of these repolarisations and possibly include these in the model of single cell motility with a Poissonian distribution. Secondly, the value of predicted noise is variable depending on the window over which persistence is measured. As with cell speed, it may be useful to employ a Bayesian inference method to approximate these parameters in future work (Rosser *et al.*, 2013; Metzner *et al.*, 2015). The higher the persistence window used, the higher the value of noise predicted. This is in part due to the cell's random repolarisations, which are not explicitly considered and would be interpreted as a jump with very high noise. Typically, a short time window of one hour was used in order to negate the effect of random cell repolarisations.

In addition to difficulties in fitting parameters, the model predictions are currently dependent on time step Δt. This can be seen in Figure @ and Table @ where simulations are run for different time steps: $\Delta t = 5$ minutes, 15 minutes and 30 minutes. These analyses suggest that the individual migratory noise is largely independent, but that the collision guidance term (w_c) is dependent on time step. At each time step a certain amount of flocking is implemented. Larger time steps result in fewer time steps, which leads to less flocking and therefore predictions of lower alignment. It will be important in future work to explore how model predictions change as the time step is varied.

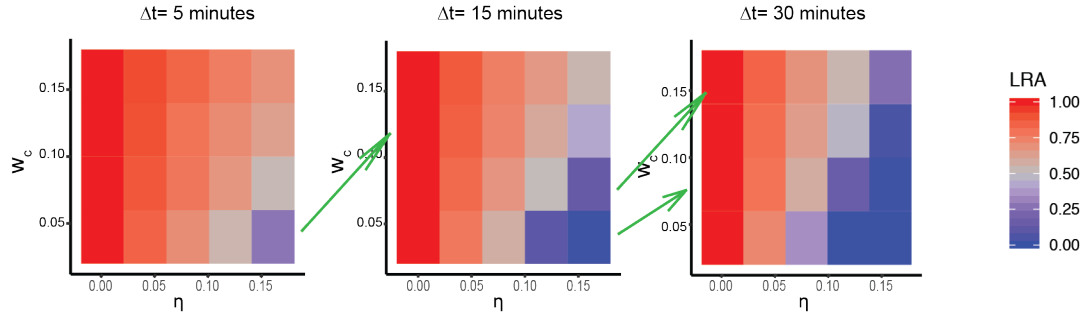


Figure 22: Heatmaps varying time step

Heatmaps of different timesteps whilst varying noise and collision guidance. N=3 simulations per point in parameter space, 700 cells are used in each simuation. Noise is scaled by Δt. Green arrows indicate simulations that have the same level of collision guidance when normalised linearly for the change in time step.

Table 4: Parameter values for testing different time steps

	$\Delta t = 5\ minutes$	$\Delta t = 15\ minutes$	$\Delta t = 30\ minutes$
Number timesteps	$t_{max} = 2100$	$t_{max} = 700$	$t_{max} = 350$
$\eta_i(\Delta t)$	$\sim N(0, \frac{\eta}{\sqrt{3}})$	$\sim N(0, \eta)$	$\sim N(0, \eta\sqrt{2})$

Overall however, fitting the parameters according to the above methods is suitable to describe the overall macroscopic behaviour when looking at the emerging patterns coordinated by hundreds or thousands of cells. By fitting noise and contact guidance of the VCAF8 aligning CAFs, the model is well able to recapitulate the emergent alignment (Figure 23). Further, this method of fitting noise can be used as a comparison between different experimental perturbations, identifying changes in noise between assays and different CAFs.

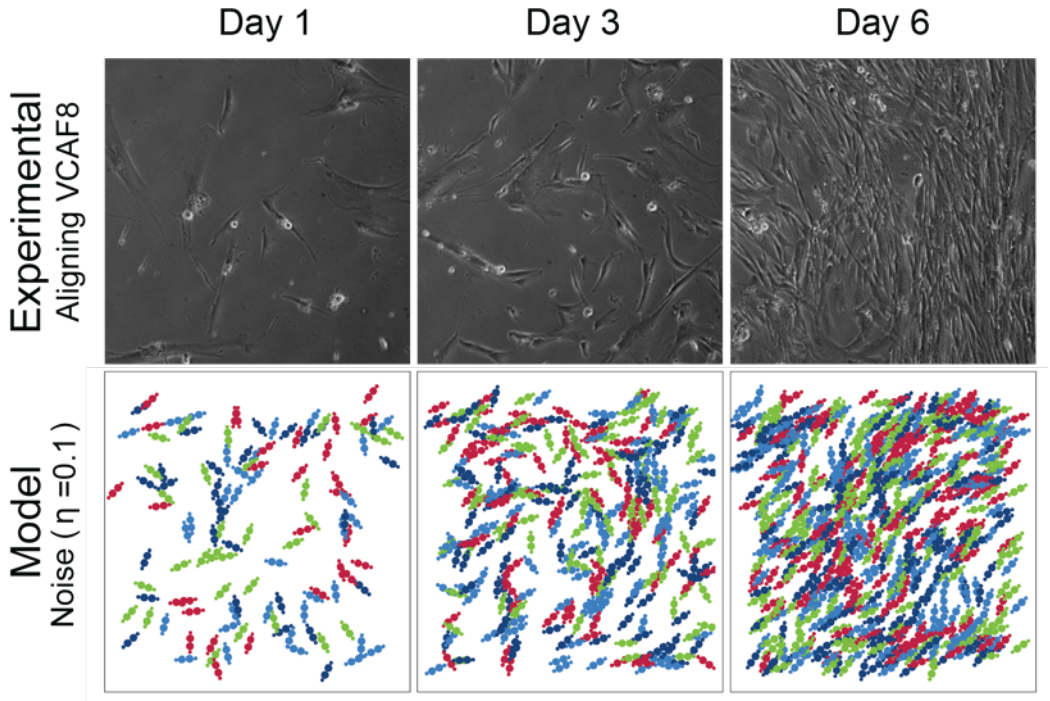


Figure 23: Model simulation versus *in vitro* experiments

Comparison of experimental alignment against model simulation with experimental persistence of VCAF8 ($\eta = 0.1, w_c = 0.04, d = 120$, all other parameters can be found in Table 3). Upper panel: bright field time-lapse microscopy of aligning VCAF8. Lower panel: model simulation of cell body organisation.

2.7 Results from the model

2.7.1 Persistence alone is insufficient to generate alignment

The persistence of the aligning and non-aligning CAFs was fitted to the noise parameter in the simulations ($Var(\eta) = 0.1, Var(\eta) = 0.18$ respectively). Even cells that are entirely persistent ($Var(\eta) = 0$) do not align without contact guidance ($w_c = 0$) (Figure 24). This suggests that variation in persistence alone is insufficient to drive alignment; additional mechanisms are necessarily required.

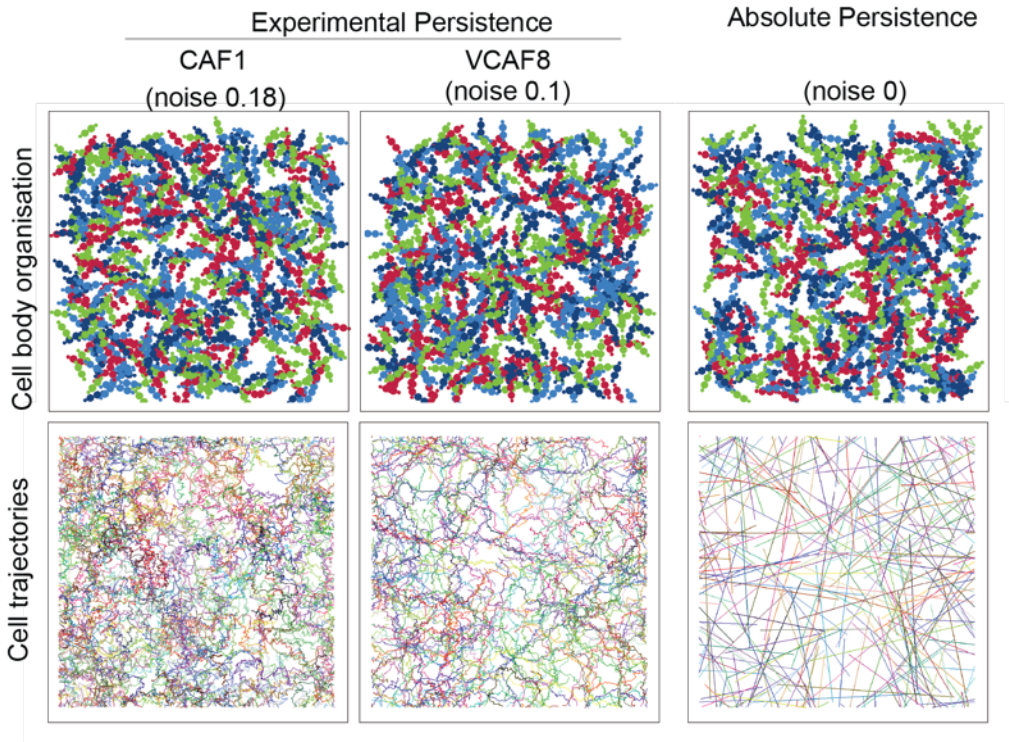


Figure 24: Cell trajectories of simulations

Mathematical model of fibroblasts (top panel) and cell trajectories (bottom panel) organisation varying cell individual migratory noise with $w_c = 0, \rho = 800, d = 0$. Panels show persistence fitted to experimental data ($\eta = 0.1$, aligned VCAF8, $\eta = 0.18$, non-aligned CAF1) as well as absolute persistence ($\eta = 0$).

2.7.2 Contact inhibition of locomotion alone is insufficient to generate alignment

To test if the introduction of contact inhibition of locomotion (CIL) could generate alignment, the model was modified so that upon collision a cell would reorient away from its neighbour by a fixed angle. The collision guidance term was set to zero ($w_c = 0$). Mathematically, the X_p and Y_p of the model could be redefined here as:

$$X_p = \cos(\theta_i(t) + \eta_i(t) \pm CIL),$$

$$Y_p = \sin(\theta_i(t) + \eta_i(t) \pm CIL),$$

where CIL was a fixed value and added or subtracted so that the cell would rotate clockwise or anticlockwise depending on which direction would reduce the impact of collision (Figure 25). In this way collision guidance and CIL are considered as separate mechanisms: collision guidance causes a continual flocking behaviour, where two cells will keep becoming more aligned at each time step so long as

contact as maintained, whereas CIL is a one-time only event that causes a cell to alter its orientation instantaneously upon contact with another cell.

Figure 26 shows that only a low level of alignment could be generated in highly contrived and unrealistic scenarios in which CIL was narrowly restricted between 10-20° and noise completely removed from the system ($Var(\eta) = 0$). Crucially for the level of persistence displayed by aligning cells experimentally ($Var(\eta)$=0.1), CIL alone was insufficient to generate anistropic patterns.

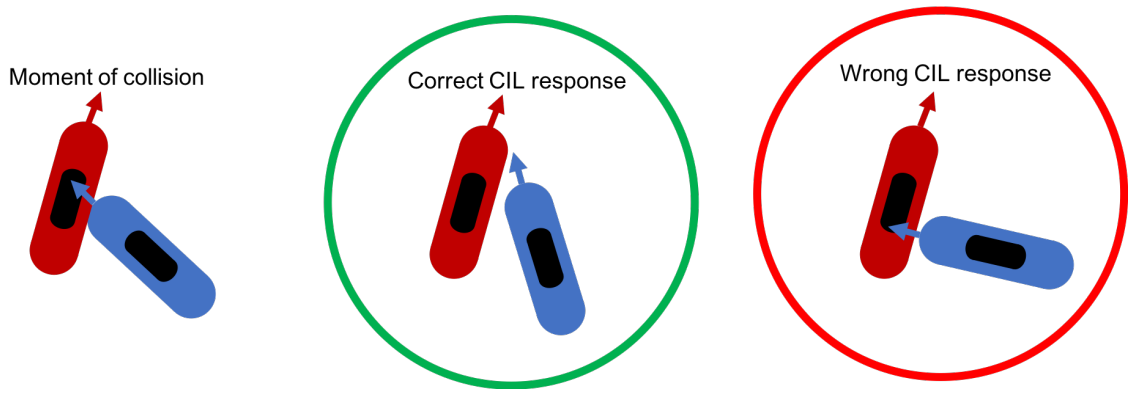


Figure 25: Schematic of cell collision response

Schematic showing how a cell repolarises upon collision to avoid becoming the cells become more overlapped. In the green circle (middle), after collision the blue cell rotates clockwise so that its direction of motion is less overlapped with the direction of motion of the red cell. In the red circle on the right, the blue cell has rotated anticlockwise so that its direction of motion will cause it to be more overlapping with the red cell.

2.7.3 The basis for a collision guidance mechanism

Coupling together these conclusions from the model together with collision guidance data from experiments (Figure 13 and Figure 14), several observations can be made. Firstly, CAFs can aptly be described as self-propelled particles (Grégoire, Chaté and Tu, 2003; Grossman, Aranson and Ben Jacob, 2008) that are not only active in generating their migration but also active in their responses to collisions. Secondly, cells involved in collisions do not bounce off each other like tennis balls nor do they align as matches shaken together in a box but rather they generate an orchestrated repolarisation that can often lead to alignment, which we define collision guidance.

Work by Grossman et al. (Grossman, Aranson and Ben Jacob, 2008) established what the minimal conditions for alignment are through inelastic collisions between self-propelled particles. They find that alignment can be generated through this passive mechanism, provided a system is in high density and particles have very high persistence. By contrast, the model presented here has much lower dependency on cell density and can encapsulate more realistic cell persistence and collision behaviours. Similarly, the work of Peruani et al. (Peruani, Deutsch and Bär, 2006) presents a force-based model of particles with inelastic collisions to study alignment. However, work from the Tumour Cell Biology group (Park *et al.*, 2019) has shown there is no correlation between cell intrinsic actomyosin activity, which is strongly linked to cell stiffness, and aligning behaviour, further suggesting that cell-cell collision responses are active. This is additionally supported by the results shown in Figure 15, whereby the faster cell involved in a collision typically undergoes more reorientation that the slower cell. This evidence serves as justification for the flocking mechanism of the model and the collision guidance weighting (w_c).

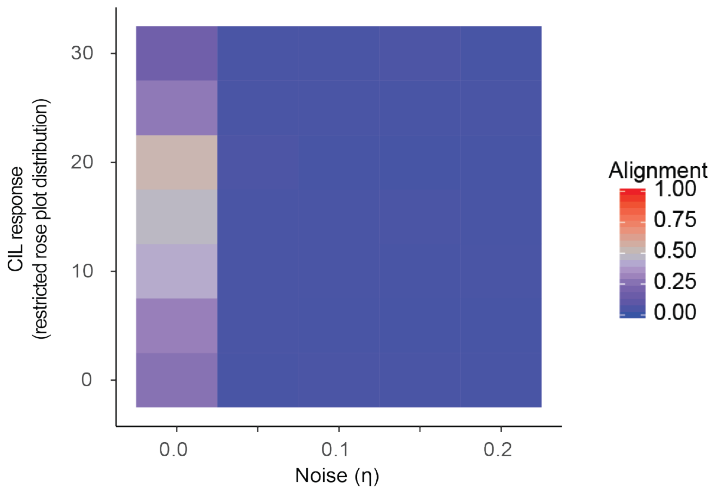


Figure 26: Exploring the effect of CIL

Model exploration of long-range alignment ($r_{min} = 0\mu m, r_{max} = 400\mu m$) as CIL response and noise (persistence) are co-varied with zero collision guidance ($w_c = 0$). All other parameters were set to default (Table 3). At noise values fitted to experimental data for aligning (VCAF8: $\eta = 0.1$) and non-aligning (CAF1: $\eta = 0.18$)

fibroblasts, there is no alignment. Alignment can only be generated when CIL is narrowly restricted between 10-20° and noise is completely removed from the system. CIL restriction is insufficient to induce any order for the range of persistence displayed by cells experimentally. Each square is the average of 10 independent simulations.

2.7.4 Exploring parameter space

Figure 27 examines how persistence (η) and collision guidance (w_c) affect global and local alignment over time, when population density is constant. It was possible to explore parameter space at a fine-grained level, since collision guidance was bound between zero and one and simulations were quick to run in parallel. Therefore, the parameters could be increased by small increments. Alignment is shown at both local (radius=0-50μm) and global neighbourhoods (radius=150-200μm) and highlights how alignment is stronger at a local level. The system can vary from complete disorder to high alignment just by varying cell

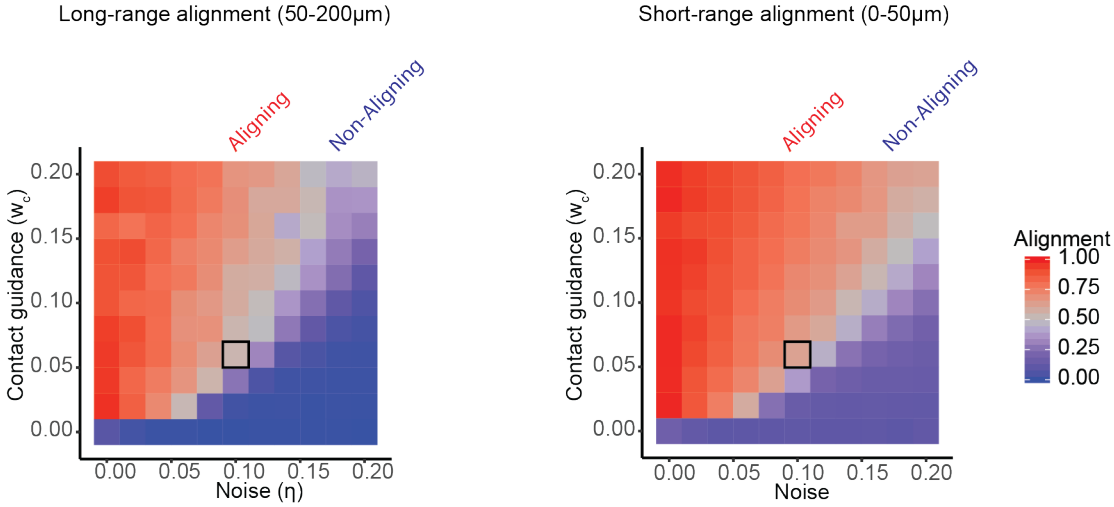


Figure 27: Model exploration of long-range and short-range alignment as noise and collision guidance are co-varied.

Heatmaps indicating levels of long-range alignment (left) and short-range alignment (right) for simulations with different parameter values for noise (η) and cell-cell collision guidance (w_c). Text above heatmaps indicates noise values fitted to experimental data for aligning (VCAF8,$\eta = 0.1$) and non-aligning (CAF1,$\eta = 0.18$) fibroblasts. For the persistence demonstrated by VCAF8 experimentally, alignment increases with increasing collision guidance. The fitted level of noise and collision guidance of the aligning VCAF8 fibroblasts is indicated on both heatmaps by a black square. Each square is the average of 10 independent simulations.

persistence. This is consistent with the variation of noise in the Vicsek model (Vicsek *et al.*, 1995), in which a phase transition occurs between order and disorder. These results indicate that both parameters are important in generating alignment, but if the cells have low enough persistence, then long range order cannot be achieved no matter the strength of collision guidance and therefore that persistence is the dominant mechanism in enabling alignment. For cells with mid-range persistence, collision guidance can determine whether the CAFs will align.

The model suggests that experimentally, the aligning CAFs with their level of persistence fall into this latter category and that the non-aligning CAFs have a level persistence beyond the threshold at which alignment can occur, regardless of the level of collision guidance. Importantly, this suggests that the lower migratory persistence of non-aligning CAFs is by itself a sufficient difference to prevent alignment. The aligning CAFs fall into the region of parameter space where contact guidance needs to be sufficiently high for alignment to occur ($c \geq 0.003$) (indicated by black square in Figure 27). It is the enhanced persistence of the aligning CAFs which facilitates the possibility of alignment however contact guidance must also be sufficiently high in these cells.

The model predicts that persistence is an influential mechanism in enabling alignment. Cells which are less polarised and therefore less persistent can override any long-term patterns of alignment that would otherwise have emerged as a result of cell-cell interactions. If there is insufficient persistence, over short periods of time, cell orientation is largely randomised. Therefore, perturbing migratory persistence in aligning fibroblasts should be sufficient to disrupt order. This was then tested experimentally. Platelet-derived growth factor (PDGF) is a potent fibroblast chemotactic that induces migratory persistence under a gradient (Seppa *et al.*, 1982). By applying an inhibitor of the receptor on fibroblasts (PDGFR) there was a statistically clear drop in persistence (**Error! Reference source not found.**), corresponding with a disruption in cell body alignment. This data, together with the model highlights the essential role of fibroblast persistence in alignment.

When noise is equal to zero, varying contact guidance can speed up or slow down alignment, but even for low values (e.g. c = 0.001), alignment will eventually occur (Figure 27). Over a mid-range of cell persistence levels, cell contact guidance becomes the determining factor in determining alignment. As described above, cells with low persistence cannot align even with strong contact guidance.

The model suggests that alignment can be achieved with sufficiently low individual migratory noise, sufficiently high cell-cell collision guidance and a sufficiently high level of cell confluence. In particular, these results indicate that alignment can occur without any specific interactions between fibroblasts and the matrix fibres they produce and that cell-cell interactions alone are sufficient to induce order. In order to test experimentally if matrix feedback was dispensable for alignment, as the model predicted, fibroblasts that generate an aligned matrix were treated with the engineered peptide `functional upstream domain' (FUD) of Streptococcus pyogenes. FUD prevents effective matrix assembly from soluble fibronectin into insoluble fibrils (Baneyx and Vogel, 1999). Cell migration is instructed by this fibril form and not the soluble form. In this way treatment with FUD precludes the cells from being guided by the matrix and we are able to eliminate matrix feedback (Figure 28A). The addition of FUD to fibroblast cultures efficiently prevented matrix fibre bundling but did not alter the migratory persistence of fibroblasts (Figure 28B) or prevent the progressive alignment of fibroblasts over time (Figure 28C). Cells are still able to align even if fibronectin deposition is blocked, confirming the model prediction that matrix feedback is not a necessary requirement for emergent matrix alignment. On the other hand

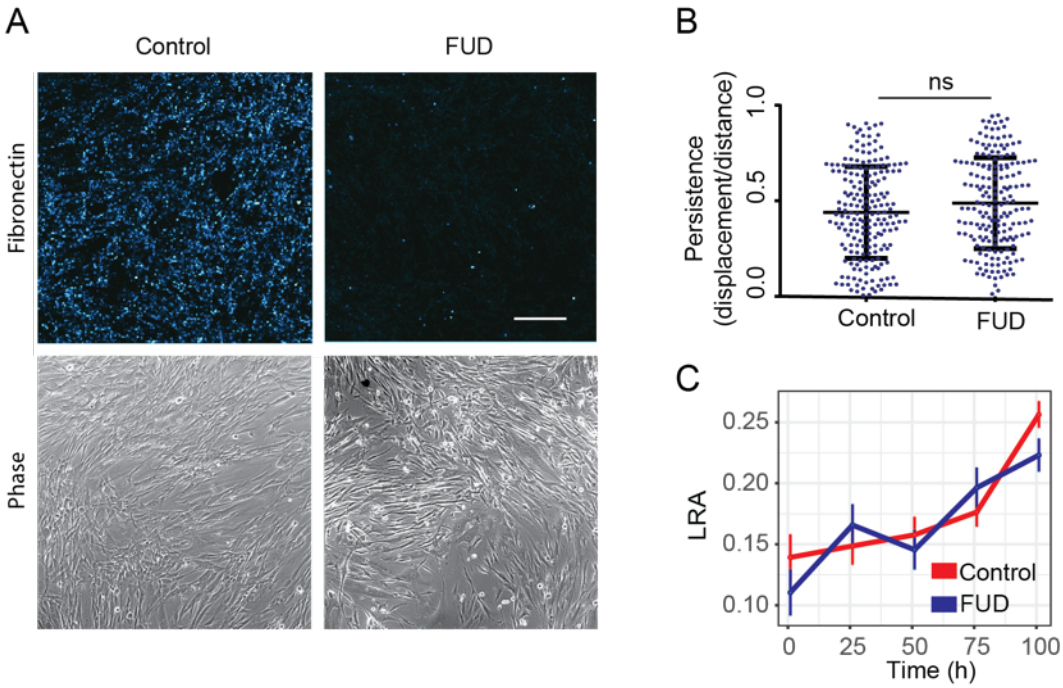


Figure 28: Cells align without matrix

Immunofluorescence of fibronectin deposited by aligning fibroblasts in the presence or absence of a Fibronectin blocking peptide (FUD). FUD blocks ECM formation. Scale bars indicate 500μm. (E) Migratory persistence of aligning cells in the presence or absence of FUD (taken over 16-hour intervals). (F) Long-range alignment over time of fibroblasts in the presence or absence of FUD. Inhibition of ECM formation by FUD treatment did not change rate of alignment.

Together, these results highlight an important interplay between the single cell (via persistence) and collective (via contact guidance) behaviours. Even though persistence is a key determinant of alignment, it must be accompanied by some degree of collision guidance. For a range of moderate cell persistence, collision guidance becomes the deciding factor.

2.7.5 Alignment requires a critical threshold cell density

The model was used to explore how alignment varies as a function of cell population size and shows that for widespread alignment to occur, a critical threshold cell density is required. Simulations were run with doubling rate of the aligning CAFs (20 hours), low noise ($\eta = 0.04$) and with high collision guidance ($w_c = 0.009$). It was found that alignment begins to occur at approximately 17% confluence. Below this level of confluence, cells do not interact enough to produce strong collective behaviour. This implies that slowing down cell population doubling

rate will perturb the rate of cellular alignment. This result is supported by data from experiments which shows that alignment only occurs locally and transiently in low confluence and further, in other flocking models (Chaté *et al.*, 2008), where particle density has to be above a certain level for alignment to occur. Cell density is therefore an important consideration when ascertaining alignment in the tumour. Lower cell density will lead to fewer cell-cell contacts and therefore collision guidance will occur less frequently.

2.7.6 The model accurately recapitulates alignment in confined spaces

So far the model has been used to explore the collective behaviour of cells in unconfined spaces, with periodic boundary conditions. However, it was of interest to test if the model could recapitulate known behaviour in confined environments. Work from the Silberzan group (Duclos *et al.*, 2014) has investigated the emergence of fibroblast alignment within confined regions *in vivo.* They found that cells confined within a strip spontaneously aligned with the confining walls. This alignment propagates through the system, away from the walls. For strips of width less than the orientation correlation length, near perfect alignment was achieved through the strip, whereas this alignment decreased towards the middle of wider strips (see Figure 29c, reproduced from (Duclos *et al.*, 2014)).

To see if the model could predict similar behaviour to the *in vitro* experiments in this new scenario, the system was adapted to include two walls a specified distance apart to create a strip. The number of cells in each strip was adjusted so that a constant high confluence would be the same in each strip. Periodic boundary conditions were still employed at the two non-confined ends of the simulation. In order to replicate the *in vivo* experiments, the volume exclusion term in the model was set to be stricter, to prevent cells going through the walls. Upon overlap with an object, cell speed was reduced to 1% of its original fixed speed until the cells were no longer overlapping. Whereas in section 2.5.2 the updating of cell positions was described by:

$$\begin{bmatrix} x_i(t+\Delta t) \\ y_i(t+\Delta t) \end{bmatrix} = \begin{bmatrix} x_i(t) + s_i \Delta t \cos\big(\theta_i(t+\Delta t)\big)\, v_e \\ y_i(t) + s_i \Delta t \sin\big(\theta_i(t+\Delta t)\big)\, v_e \end{bmatrix}$$

where v_e is a proxy for volume exclusion defined by

$$v_e = \begin{cases} 0.25 \text{ if the head node of a cell is overlapping with any other node of another cell,} \\ 1 \qquad \text{otherwise} \end{cases}$$

instead v_e was defined as:

$$v_e = \begin{cases} 0.01 \text{ if the head node of a cell is overlapping with any other node of another cell,} \\ 1 \qquad \text{otherwise} \end{cases}$$

Finally, Duclos et al. observed that a fibroblast near the wall will align with the wall even at low confluence. To account for this in the model, collision guidance was activated in the fibroblasts upon contact with the walls.

Simulations were run at high confluence with strip widths of $200\mu m, 600\mu m, 1000\mu m$ and two different levels of $\eta = \{0, 0.1\}$. The results from the model do indeed agree with those of Duclos et al (Figure 29). For a narrow strip of $200\mu m$, nearly total alignment is achieved as the alignment with the walls is able to propagate through the whole system. For a strip width of $600\mu m$, alignment towards the centre of the strip decreases. Interestingly, this effect is much stronger in cells with high noise, since the propagation of alignment to the wall is lost much more quickly through the system. Finally, for a large strip width of $1000\mu m$, overall alignment is low as the majority of cells are not close enough to the walls to align with them. Overall, it is promising that the model can recapitulate these experimental results and could therefore be used to answer more precise questions about this experiment eg critical strip width for alignment and dominant effects of more complex confining regions.

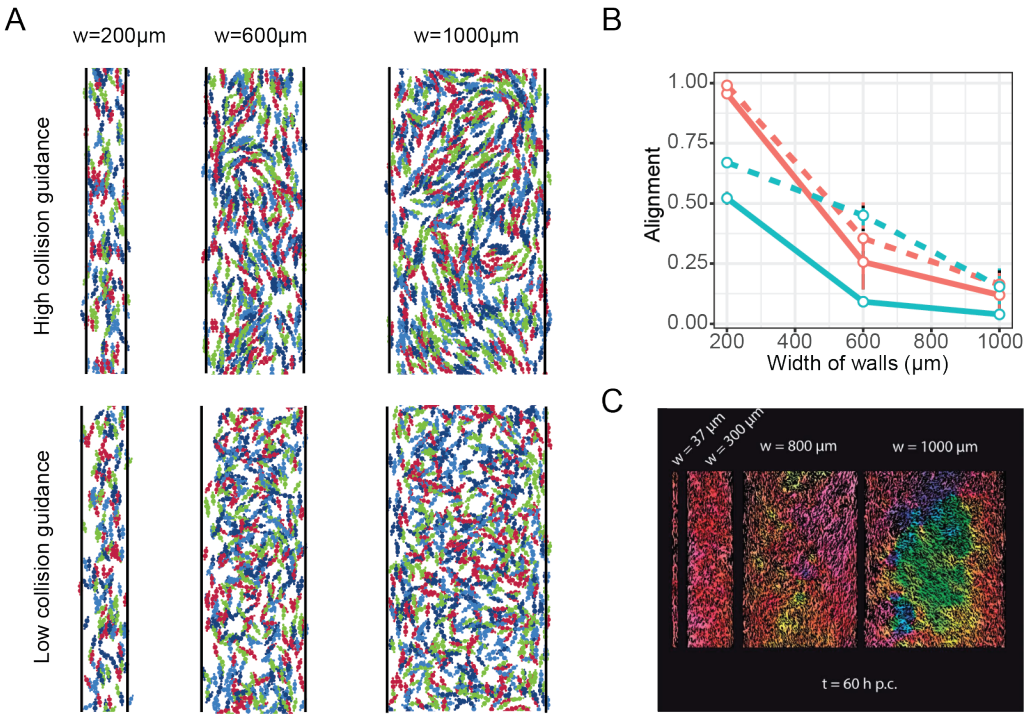


Figure 29: Emergent behaviour in confined spaces

The model was challenged to reproduce the width-dependence findings by Duclos et al. (A) Stills from simulations at high confluence with width between walls of $200\mu m, 600\mu m, 1000\mu m$ respectively and $\rho = (200,600,1000)$ respectively, noise (η)=0.1 for high collision guidance ($w_c = 0.05$) and low collision guidance ($w_c = 0.01$). (B) Graph showing alignment of simulations run varying width between walls, for low noise ($\eta = 0$, orange) and high noise ($\eta = 0.1$, blue), and for for low collision guidance ($w_c = 0.01$, solid line) and high collision guidance ($w_c = 0.05$, dashed line). Simulations were run at high confluence, simulating emergent behaviour after seven days. N=5 simulations per point in parameter space. (C) Image reproduced from (Duclos *et al.*, 2014) shows agreement between model and experimentation.

2.7.7 Predicting emergent behaviour in heterogeneous populations

Heterogeneity not only exists within the tumour, but also within the stromal CAF population. Indeed multiple sub-populations of CAFs have been identified with different functional roles (Augsten, 2014; Kalluri, 2016; Öhlund *et al.*, 2017). It is important to understand how heterogeneity in CAFs might affect alignment, and whether suppressing or enhancing heterogeneity could have potential therapeutic applications. The model was used to explore the effects of increasing heterogeneity and to predict the circumstances under which alignment might be disrupted.

A first level of heterogeneity was simulated – in which there is heterogeneity of persistence in a single population. Indeed, experimental data shows a small subpopulation within the aligning category with persistence values close to the mean of non-aligning CAFs (Figure 30A). We explored whether this level of heterogeneity might compromise alignment. Simulations were run where each cell i had its own unique persistence $Var(\eta_i)$ drawn from the experimental distribution, varying only collision guidance. With this level of heterogeneity cells were still able to align with moderate collision guidance.

Next, simulations were run with the level of noise and collision guidance to match the aligning CAFs experimentally but with a user-defined probability that upon collision, a cell would reorient with a completely random uniformly distributed orientation. This mimicked the CIL response largely seen in the non-aligning CAFs. The model predicts that alignment is quickly disrupted by adding in as few as 10% non-aligning CAFs (Figure 30B and C) and therefore that aligning populations of cells must necessarily dominate for alignment to occur.

This prediction was then tested experimentally. Aligning and non-aligning cells were labelled with membrane localized fluorophores and seeded at different starting proportions. Co-cultures were imaged every second day so as to measure cell body organisation over time (Figure 31). With the addition of as little as 20% non-aligning CAFs, alignment was almost entirely destroyed, confirming the model predictions.

Taking the computational and experimental results together strongly points to the fragility of alignment and its sensitivity to heterogeneity. This raises an interesting question as to whether non-aligning cells all exhibit non-aligning characteristics or whether they possess a subgroup of cells with aligning characteristics that are simply unable to over-ride the dominant non-aligning phenotype. Most importantly, these results draw attention to the efficacy of altering the balance in a heterogeneous mixture of CAFs as a means of destroying alignment and the potential power of using this therapeutically. This offers a mechanistic view on how limited alignment in a highly heterogeneous tumour micro-environment can be achieved. Given the fragility of emergent alignment, for an area of alignment to

occur, the population would need to be largely dominated by CAFs displaying an aligning phenotype. It can be conjectured that this sensitivity of alignment applies *in vivo* and therefore only small phenotypic perturbations might be required to dissolve harmful microstates of alignment.

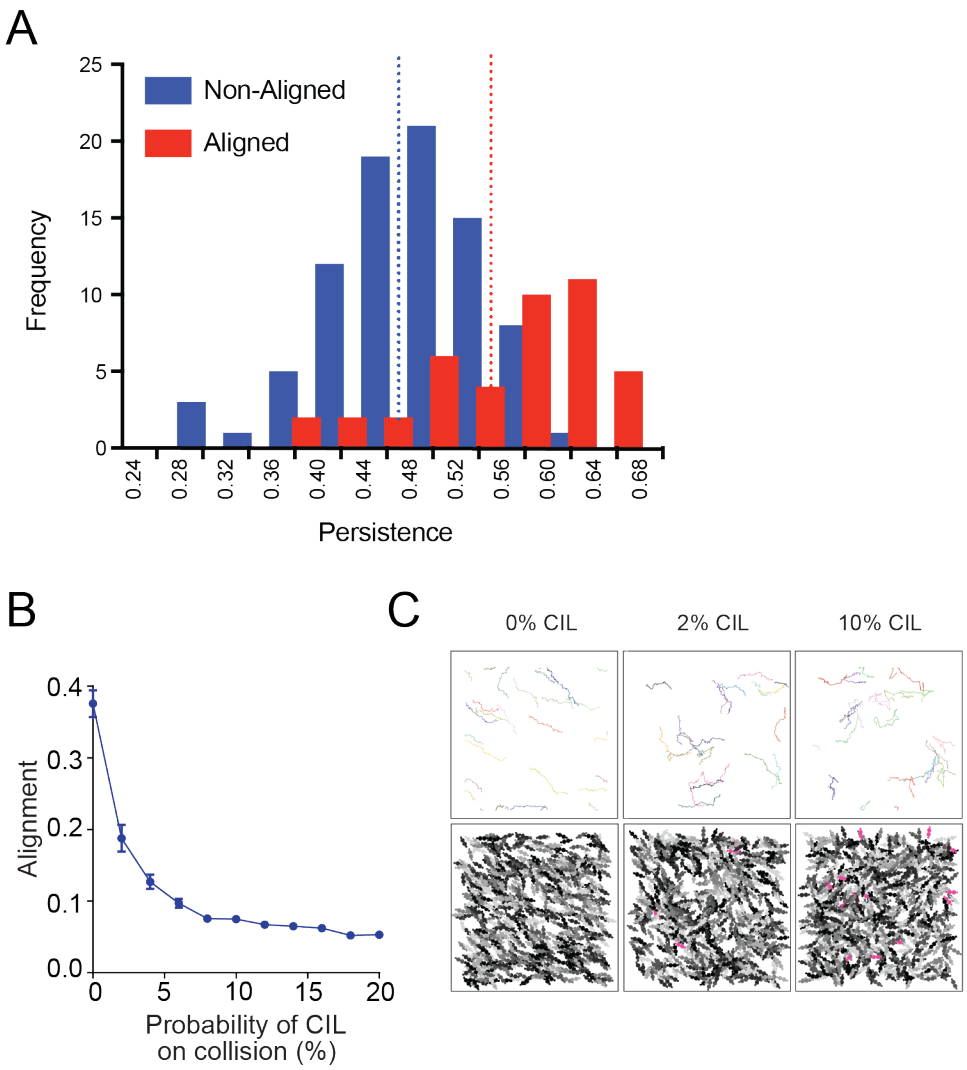


Figure 30: Exploring alignment in heterogeneous environments.

(A) Histogram of the persistence distribution in aligning (red) and non-aligning (blue) populations for distances travelled over 0.5h. The dotted line indicates the mean of both populations. (B) Model simulation of cell organisation with increasing probability of CIL upon collision, starting from a system that aligns ($\eta = 0.1, w_c = 0.06$). CIL angle is drawn from the experimental data of non-aligning cells in Figure 13. Increasing CIL response is sufficient to disrupt alignment. Each value is the mean and SEM of 10 independent simulations. (C) Representative simulation of migration paths (upper panel) and cell body organisation (lower panel) with an increasing probability of CIL on collision. CIL events are shown in pink.

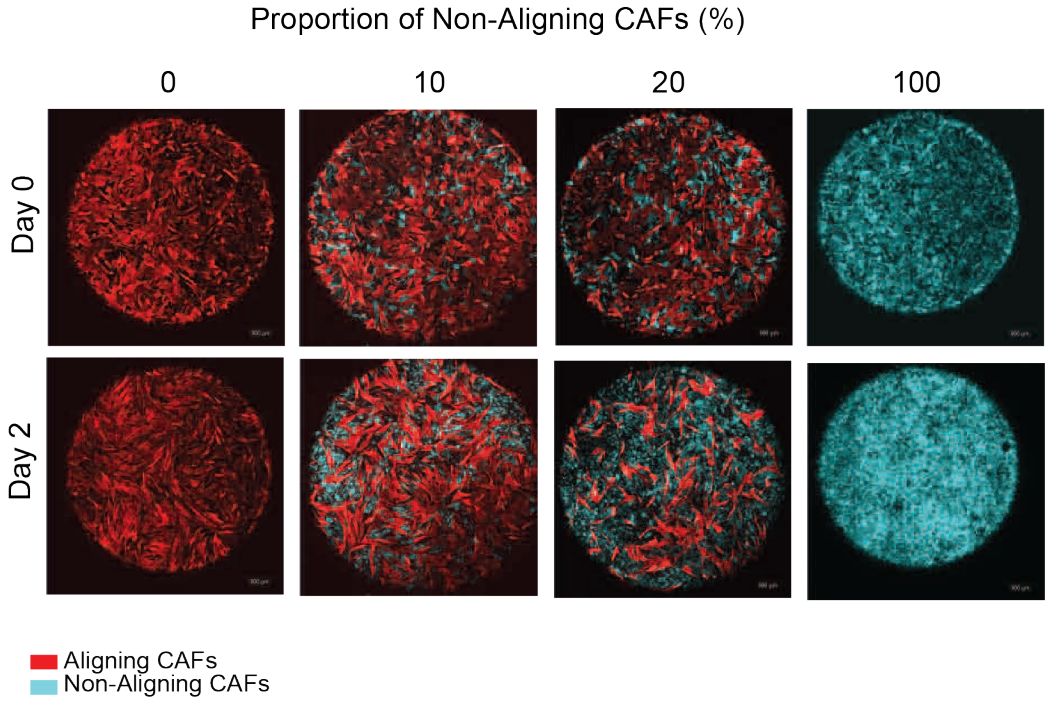


Figure 31: Co-culture experiments *in vitro*

Experimental exploration of alignment with increasing heterogeneity. Aligning (red) and non-aligning (cyan) fibroblasts were mixed at the proportions indicated and imaged every second day.

2.8 Chapter highlights

A summary of the main findings of this chapter is given as follows:

- Aligning fibroblasts have enhanced individual migratory persistence
- High persistence alone is insufficient to generate alignment
- Fibroblasts align through an active collision guidance mechanism
- Alignment is a fragile state that can be destroyed by heterogeneity

The model predicts that cell-cell collision guidance is a requirement for alignment to emerge. As a direct consequence of this finding, Danielle Park in the Tumour Cell Biology laboratory was able to identify a transcription factor TFAP2C, upstream of RhoE, thereby influencing actomyosin contractility, that was elevated in aligning fibroblasts (Park *et al.*, 2019). Upon collision, aligning fibroblasts displayed actomyosin contractility at the point of contact, causing cells to reorient to each other and consequently align, as predicted by the collision guidance mechanism in

the model. In this way, a combination of computation and experimental investigation has helped to elucidate emergent cell patterning and how harmful aligning patterns might be disrupted therapeutically.

Chapter 3. Modelling fibroblast-matrix interactions

In the previous chapter, it was shown how different mechanistic properties relating to the motility of fibroblasts were responsible for producing different collective behaviours: isotropy and anisotropy (Figure 2). This in turn produced a similarly patterned extracellular matrix. Fibroblasts displayed key differences in their migrational persistence and in their interactions with each other upon contact, which could manifest as collision guidance or contact inhibition of locomotion. However, a third piece of this puzzle has remained unaddressed: namely, what is the role of the extracellular matrix upon the emergent behaviour of fibroblasts and hence on the matrix itself? Fibroblasts are largely responsible for the deposition, rearrangement and degradation of ECM fibres, but could these fibres conversely and simultaneously exert a role on fibroblast behaviour? If so, what would the emergent consequences of such a feedback loop between fibroblasts and ECM be?

3.1 Motivated by experiments

3.1.1 Diverse matrix patterns are found *in vivo* and can be quantified

The ECM plays a crucial role in tissue function. A large part of this is due to the topology and physical organisation of the matrix for example, highly aligned linear bundles of ECM are found in the tendon, enabling the extension and contraction of limbs. To see this diversity of organisation in non-pathological tissues, second harmonic imaging of collagen fibres *in vivo* was carried out using murine organs (Figure 32A). The position of fibroblasts was established using mice with transgenic expression of the fluorescent nuclear marker, H2B-GFP in fibroblasts. The long axis of the fibroblast nuclei indicates the orientation of the cell body and is correlated with the fibres they deposit and confirms this assumption in the model (Figure 5 from previous chapter). The dermis ECM is largely made up of curved bundles of fibres, giving the skin its structural integrity. The porous structure of the liver could possibly facilitate transportation of materials in and out.

In order to formally describe the matrix patterns quantitatively it was necessary to introduce several metrics. Five metrics were derived: long-range alignment (LRA), short-range alignment (SRA), percentage of high-density matrix (HDM), curvature (Curv) and fractal dimension (Frac). These metrics together could characterise different matrix properties. For example, a 'swirl-like' matrix pattern maintains high short-range alignment as neighbouring fibres follow largely similar directions but low long-range alignment (dermis, Figure 32A). Such a matrix would also have high curvature.

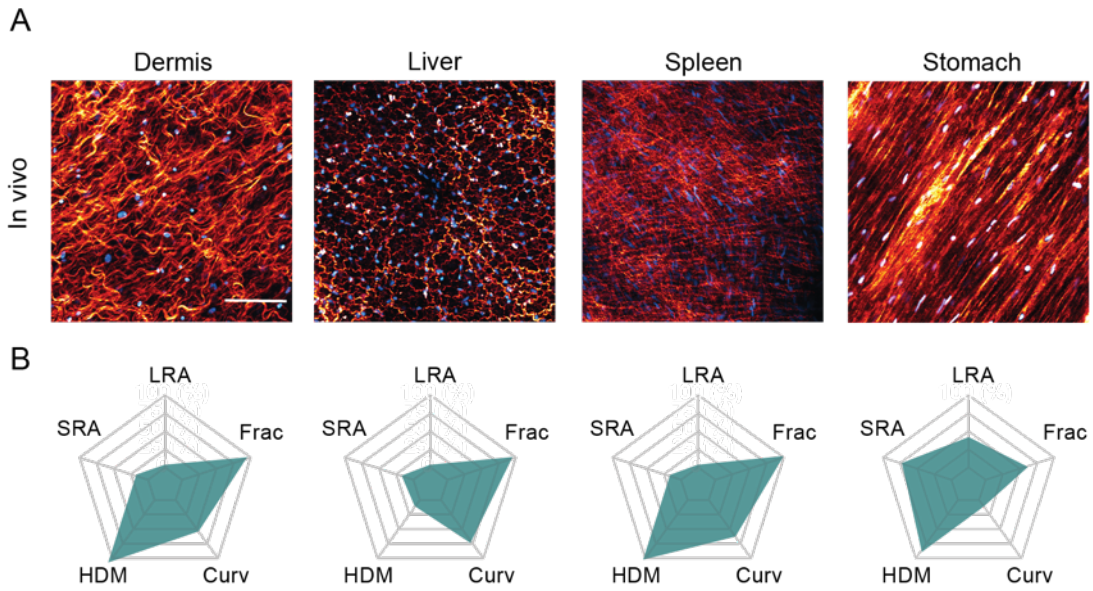


Figure 32: Diversity of mouse tissues

Second harmonic imaging of collagen (orange) in various tissues from a PDGFR nuclear labelled mouse. PDGFR positive nuclei (cyan) indicates fibroblast density. (B) Each matrix is characterised on a starplot below the image according to long-range alignment (LRA), short-range alignment (SRA), percentage of high-density matrix (HDM), curvature (Curv) and fractal dimension (Frac). Scale bars represent $100\mu m$.

3.1.2 Metrics describing the matrix

Long and short-range alignment (LRA, SRA)

The method of measuring fibre alignment is the same as for measuring cell bodies as described in section 2.2. Fibre position and orientation of the *in vivo* images was

established using CT-Fire (Bredfeldt *et al.*, 2014) by Robert Jenkins from the Tumour Cell Biology Laboratory.

CT-Fire was used to extract individual fibres from 512x512 greyscale images. These were then fed into a MATLAB script, which computed the orientation of each fibre assuming that it was straight between its start and end point. These angles were then input into a 512x512 mathematical matrix. Interpolation was applied to fill in empty matrix values and followed by a smoothing mean filter. The median difference in angle was then computed for increasing radii. In this work, values at distances of $100\mu m$ and $200\mu m$ are used. The final output for a given radius was the sine of the median angle difference. At all steps, care was taken to respect the periodicity between angles, for example, two fibres orientated at $175°$ and $5°$ respectively have an angle difference of $10°$, not $170°$ (Park *et al.*, 2019). Figure 8 shows a schematic of how alignment is computed.

Percentage of high-density matrix (HDM)

The percentage of high-density matrix indicates the spatial heterogeneity of the distribution of fibres. If cells are channelled to cover specific regions, this corralling behaviour would lead to some of areas of high-density matrix, and also to areas of very low-density matrix where there has been very little cell coverage (liver, Figure 32). A schematic of this is given in Figure 33. Limitations of this method lie in the subjectivity of thresholding value beyond which an area is considered to have high-density matrix. Other methods which circumvent this problem include Ripley's K function and Moran's I for spatial autocorrelation (Nawaz *et al.*, 2015; Yuan, 2016).

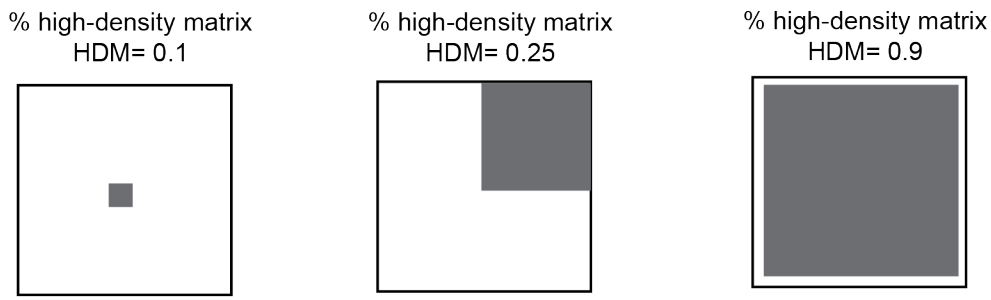


Figure 33: Schematic of High-density matrix (HDM)

High-density matrix (HDM) is a measure of the proportion of pixels in image above a user-specified threshold intensity.

Curvature (Curv)

For quantification of matrix curvature, images were split into RGB channels. The mask of the image was generated using the "Ridge detection" plug-in in ImageJ (Steger, 1998). Given a user-specified curvature window, the plug-in Anamorf (Barry, Williams and Chan, 2015) computed the mean curvature of the mask, measured as the change in angle moving incrementally along the mask lines (Figure 34). The curvature window must be chosen with care to encapsulate the rate at which lines curve. If the curvature window is too small, even tight curls seems straight and if it is too big then changes in curvature could be missed.

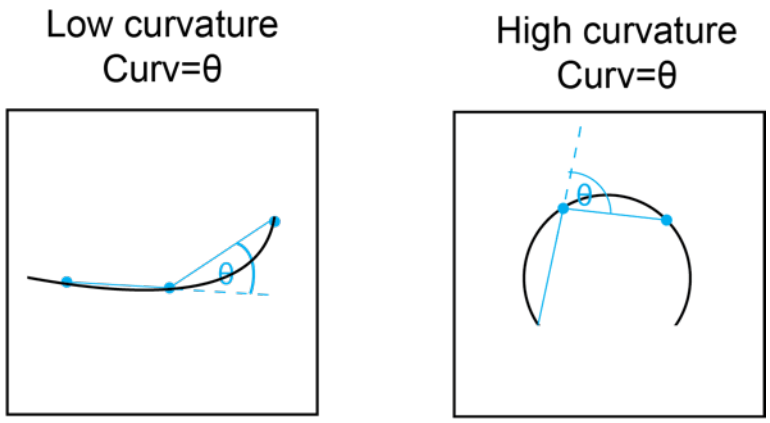


Figure 34: Schematic of curvature (Curv)

For a user-specified window, curvature of a curve is measured as the change in angle as the curve is traversed in increments of the specified window, as indicated here by θ. In an image with many curves, the curvature is an average of the curvature of all curves.

Fractal dimension (Frac)

Fractal dimension is an indicator of the self-similarity of the matrix and is bound between in the range [1,2]. A fractal dimension of 1 would represent a straight line and a fractal dimension of 2 would represent a curve covering every point in two-dimensional space, for example the limit of the Hilbert Curve (Figure 35A). Specifically, the metric used by the ImageJ plugin Anamorf(Barry, Williams and Chan, 2015) is the box-counting dimension. A grid with squares of side length ϵ is

overlaid on a mask image of a matrix pattern. The number of squares which are occupied by part of the mask N is recorded. Fractal dimension is then given as the limit of the number of grid squares required to cover the mask for increasingly smaller squares as ϵ decreases.

Other metrics that could have been included such as Getis-Ord hotspot analysis (Heindl, Nawaz and Yuan, 2015; Nawaz *et al.*, 2015; Yuan, 2016) and lacunarity. However, the chosen five metrics with user-defined length scales for long-range alignment, short-range alignment, curvature and the threshold for high-density matrix sufficiently captured key differences between matrix patterns and enabled direct comparison. Furthermore, these metrics could then be easily visualised in the form of starplots with five axes (Figure 32B).

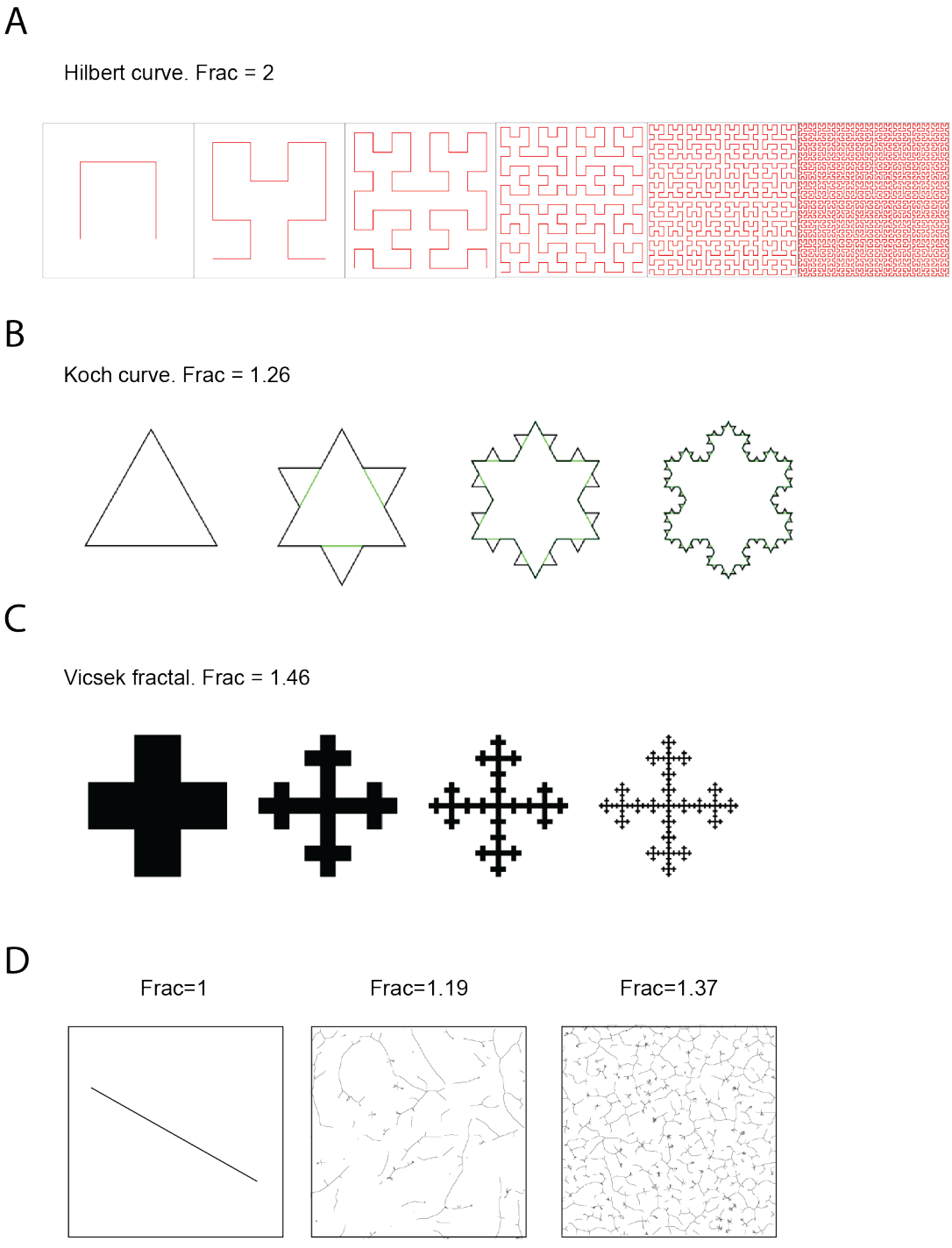


Figure 35: Examples of fractal dimension (Frac)

Fractal dimension (Frac) is a measure of how a one-dimensional curve fills two-dimensional space and can be thought of as the self-similarity of the curve. (A) The construction of the Hilbert curve, which at its limit has Frac=2. (B) The Koch curve, often known as the snowflake fractal, is the limiting shape of the sequence shown and has Frac=1.26. (C) The Vicsek fractal has Frac = 1.46. (D) Further examples of masks with different values for Frac. These masks come directly from matrix patterns generated with the model. Hilbert curve in (A) comes from

https://en.wikipedia.org/wiki/Space-filling_curve. Koch curve in (B) comes from https://en.wikipedia.org/wiki/Koch_snowflake. Vicsek fractal in (C) comes from https://en.wikipedia.org/wiki/Vicsek_fractal

3.1.3 Establishing the role of feedback between fibroblasts and ECM

In Chapter 2, it was shown that it is cell-cell interactions together with individual migratory persistence that are sufficient to generate higher-order alignment *in vitro* in the absence of matrix (Figure 28). It therefore remains to establish whether or not matrix topology impacts on cell behaviour and under what circumstances this could occur.

To investigate this directly, fibroblasts were plated on thick matrices with differing degrees of alignment, grown over seven-day assays to full confluence. It was found that these plated fibroblasts moving on a thick aligned matrix had increased persistence as compared with the same fibroblasts moving on an isotropic matrix (Figure 36: Experimental confirmation of matrix feedback, p=0.0064, two-tailed t-test). This suggests that matrix topology exerts a change in fibroblast behaviour, thereby ultimately reinforcing the matrix organisation.

It is difficult to see the real-time coevolution of fibroblasts and matrix fibres experimentally. Computational modelling can therefore be of great use in unpicking the “chicken and egg” problem of how fibroblasts and matrix interact and influence each other to generate higher-order organisation.

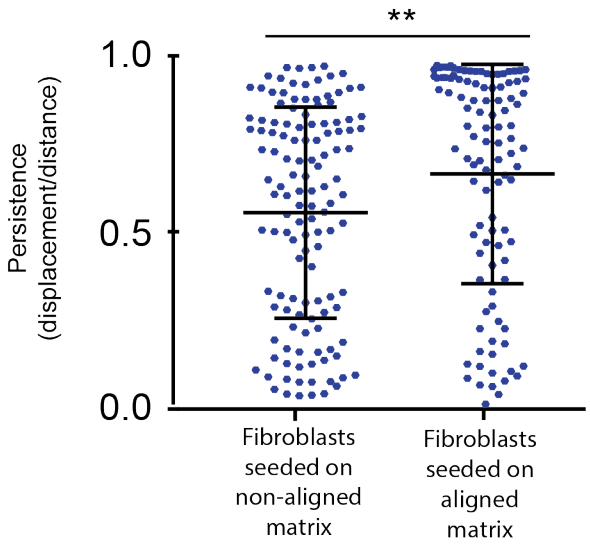


Figure 36: Experimental confirmation of matrix feedback

Migratory persistence of aligning or non-aligning fibroblasts seeded on a thick aligned matrix (over 8-hour windows). On an aligned matrix, fibroblasts migrate with higher persistence ($p = 0.0064$, using a two-tailed t-test).

3.2 Model construction II

To understand what the consequences of matrix feedback might be in the emergent patterning of fibroblasts and ECM, the Vicsek-inspired model described in the previous chapter was extended to incorporate fibroblast-matrix interactions. The model was modified to include a second layer representing the ECM. This brought the model into a pseudo-three-dimensional environment with a top layer consisting of fibroblasts and the bottom layer consisting of matrix. This ECM layer was arranged as a grid and each grid point was associated to matrix fibres that could be oriented in eight possible bins, in increments of $\frac{\pi}{8}$ within the range $[0, \pi]$, reflecting the nematic interactions of fibroblasts with fibres. In this model, fibroblasts could generate, delete or assign a different directionality to matrix fibres, representing the deposition, proteolytic degradation and rearrangement of the matrix, respectively. Additionally, fibroblasts could alter their orientation to follow matrix fibres, creating a two-way feedback between fibroblasts and the matrix fibres they produce. The main innovation of the model away from the traditional Vicsek model is in adding the ability of cells to be guided by the matrix fibres below it whilst simultaneously producing and reorganising these fibres (Figure 37: Schematic two-

layer model). The original set-up and cell shape in the model remain unchanged from 2.5.

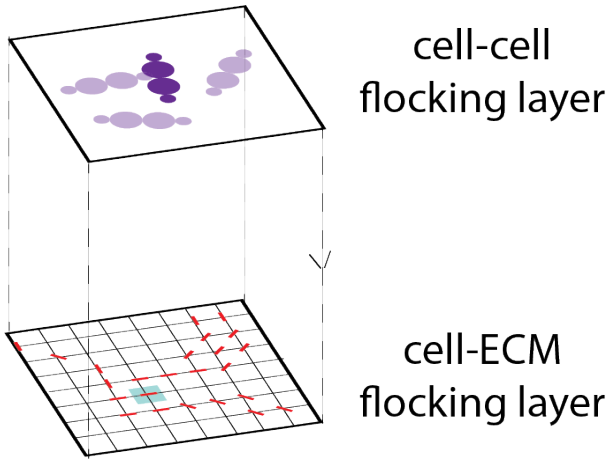


Figure 37: Schematic two-layer model

Schematic showing the two layers of the model and the interplay between them. Fibroblasts flock with each other on the top layer. Simultaneously, fibroblasts deposit matrix fibres (red) onto an underlying grid points, representing the layer of ECM. These fibres are then able to influence the orientation of the fibroblasts above through matrix feedback.

The model has several similarities with an existing model presented by Dallon et al. (Dallon, Sherratt and Maini, 1999), who model fibroblasts as discrete elements and the ECM as continuous to investigate the interplay between fibroblasts and matrix in the context of wound healing. This is an interesting and highly relevant model and would have been a reasonable choice for exploring the thesis research questions. However, the model presented here is perhaps more suitable due to the incorporation of clear rule-defined cell-cell interactions and cell persistence, which are lacking in Dallon et al. In addition, the grid point structure of the ECM acts as a hybrid, where fibres can be modelled in more detail than in the continuum (Dallon, Sherratt and Maini, 1999) whilst remaining computationally inexpensive. This means cells can interact iwht individual groups of fibres, enabling more insight into the feedback mechanism between CAFs and matrix fibres.

The fineness of the grid could be determined by the user, however by default, the number of gridpoint was determined as follows:

$$numberGridPoints = \left\lceil \frac{simulationWindowSize}{r} \right\rceil \quad (25)$$

where r is the radius of the body nodes of the cells (Table 3). This means that the distance between grid points is slightly smaller than the small head node of a cell. This grid is distinct from the voxel grid described in section 2.5.3. The cell speed is set so that a cell will move approximately half the side length of an ECM grid square. This means that trails of ECM produced by a cell will be deposited in a spatially continuous manner ie that fibres will lie in adjacent grid boxes without gaps. If time step were significantly increased, cell speed significantly increased or ECM grid made much finer, this could lead to gaps in the ECM (Dallon, Sherratt and Maini, 1999) causing the cell not interacting with fibres it otherwise would have.

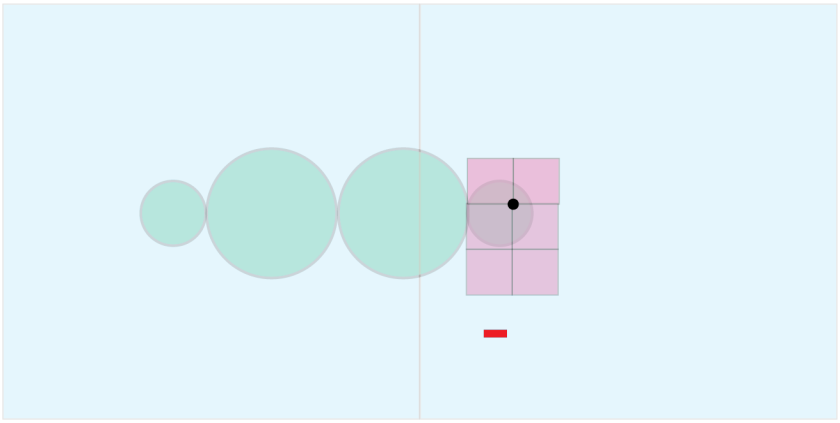


Figure 38: Schematic of ECM grid and voxel grid

The default ECM grid (pink) has squares of size slightly smaller than the cell head. The cell is associated with the ECM grid point closest to the centre of its head (indicated by a black dot). The voxel grid (blue) used for establishing cell neighbours for computational speed and cell proliferation has squares of side length slightly larger than the length of a cell. In a single time step, the cell moves approximately the distance indicated by a red line, which is slightly smaller than the length of an ECM grid square.

When establishing the model, it was of importance to investigate in detail how fibronectin is deposited by a moving fibroblast. To this end, fibroblasts were plated, then allowed to spread and commence migration. They were then fixed and stained for F-actin to reveal their dominant protrusion and hence direction of migration, paxillin to reveal points of substrate attachment, and fibronectin (Figure 39). This indicated that the formation of fibronectin puncta and fibrils at the basal surface of fibroblasts was clearly apparent ahead of the cell nucleus, most notably in the zone behind the paxillin-rich adhesions and less obviously in the leading protrusion. The formation of fibronectin into large fibrils was more prominent towards the cell rear.

These data indicate that fibronectin deposition begins in a zone well ahead of the nucleus and near cell matrix adhesions, and suggest that fibronectin becomes progressively organised into fibres toward the cell rear. Based on these findings, it seems most appropriate to have fibre deposition at one of the front two beads of the cell. This experiment was carried out by Erik Sahai and Steven Hooper from the Tumour Cell Biology laboratory.

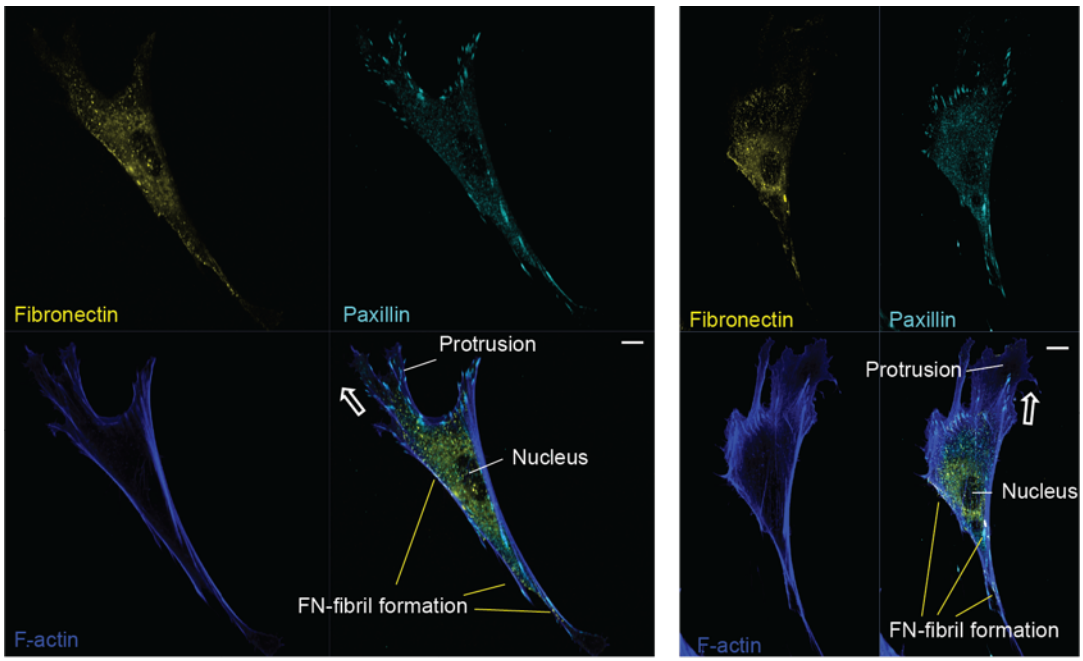


Figure 39: How fibroblasts deposit fibres

Two examples exploring where fibroblasts produce fibronectin. Fibroblasts were plated, and then began to spread and commence migration. They were then fixed and stained for F-actin (bottom left panels) to reveal their dominant protrusion and hence direction of migration, fibronectin (top left panels) and paxillin to reveal points of substrate attachment (top right panels). A composite image is shown in the bottom right panels. Scale bar represents $10\mu m$. Experiment and data collection carried out by Erik Sahai and Steven Hooper.

3.2.1 Model overview

As described in section 3.2.1, the ECM consists of a fine grid, with the distance between grid points being at least as small as a cell's head. At time t, a cell i with orientation $\theta_i(t)$ is associated with and will interact with whichever grid point k is closest to the centre of its head (Figure 38). In this way, a cell is always associated with exactly one grid point.

The organisational and matrix guidance components of the model are implemented algorithmically and are described here. The grid point k consists of eight possible bins in increments of $\frac{\pi}{8}$ within the range $[0, \pi]$. Three new organisational parameters are introduced: $o_{dep}(\Delta t)$:= number of fibres deposited by a fibroblast per time step, $o_{deg}(\Delta t)$:= number of fibres degraded from each bin by a fibroblast per time step, $o_{re}(\Delta t)$:= number of fibres rearranged from neighbouring bins by a fibroblast per time step.

The fibre organisational terms $o_{dep}(\Delta t), o_{deg}(\Delta t)$ and $o_{re}(\Delta t)$ are assumed to be linearly dependent on the time step Δt, eg the number of fibres deposited over two time points ought to be equal to the number fibres deposited in a single time point of twice the length. There are many unanswered questions regarding the way in which fibres are deposited by fibroblasts together with quantification of the organisation of the fibres. Therefore, the model is used in an exploratory manner, assuming linear deposition of fibres and where time t is measured in units of 15 minutes and Δt is varied with respect to a frame rate of one unit. In the rest of this work, the assumption is made that $\Delta t = 1$ unit ie fifteen minutes. In an abuse of notation o_{dep} is used in place of $o_{dep}(\Delta t)$, similarly for o_{deg} and o_{re}. It will be of interest in future work to investigate if the assumption of linearity is correct and how predictions vary depending on the time step. Fibroblasts deposit fibres in a discrete manner at each time step, rather than continuously. Therefore, as time step becomes large, or indeed if cell speed becomes large, eventually fibres will be deposited by a cell in non-adjacent ECM grid points, creating a dashed trail of ECM fibres. The current time step is small enough such that the cell moves less than the length of an ECM grid square (Figure 38). Therefore, whilst fibres are deposited discretely, they are deposited in adjacent ECM boxes, reminiscent of the continuous path of fibres observed experimentally.

The implementation of these parameters is best illustrated through an example:

At time t suppose the information about fibres at grid point k can be written

$$k = [1,8,3,3,0,0,0,2]$$

ie there is 1 fibre in the bin with orientation 0, 8 fibres in the bin with orientation $\frac{\pi}{8}$, 3 fibres in the bin with orientation $\frac{2\pi}{8}$ etc. The number of fibres in each bin can never hold a negative number of fibres. Suppose cell i has orientation $\theta_i(t) = \frac{\pi}{1000}$ and is associated with grid point k.

Choosing fibre for matrix guidance

Cell i will be guided by fibres in a single bin in box k. This bin is called the dominant orientation and is denoted $\phi_k(t)$. This is chosen through a Gillespie algorithm where bins are weighted according to the number of fibres in them so that cells probabilistically choose to be guided by a bin with more fibres in. In the example, the total number of fibres in grid point k is $1 + 8 + 3 + 3 + 0 + 0 + 0 + 2 = 17$. Then the dominant orientation for the cell to follow is defined by:

$$\phi_k(t) = \begin{cases} 0, \text{ with probability } \frac{1}{17} \\ \frac{\pi}{8}, \text{ with probability } \frac{8}{17} \\ \frac{2\pi}{8}, \text{ with probability } \frac{3}{17} \\ \frac{3\pi}{8}, \text{ with probability } \frac{3}{17} \\ \frac{4\pi}{8}, \text{ with probability } \frac{0}{17} \\ \frac{5\pi}{8}, \text{ with probability } \frac{0}{17} \\ \frac{6\pi}{8}, \text{ with probability } \frac{0}{17} \\ \frac{7\pi}{8}, \text{ with probability } \frac{2}{17} \end{cases}$$

Suppose in this instance that the value chose is $\phi_k(t) = \frac{\pi}{8}$. The way in which cell i is guided by this bin is described in section 3.2.4.

The cell will then reorganise the matrix fibres in grid point k. First fibres will be degraded, then rearranged, and then finally the fibroblast will deposit new fibres as described below:

Fibre degradation

All bins in grid point k are depleted at each time step by o_{deg} so that, in the example

$$k_{temp1} = (\max(0,1 - o_{deg}), \max(0,8 - o_{deg}), \max(0,3 - o_{deg}), \max(0,3 - o_{deg}),$$
$$\max(0,0 - o_{deg}), \max(0,0 - o_{deg}), \max(0,0 - o_{deg}), \max(0,2 - o_{deg}))$$

In the example where $o_{deg} = 1$, after degradation

$$k_{temp1} = (0,7,2,2,0,0,0,1)$$

Fibre rearrangement

Fibres from the two neighbouring bins will be moved to the bin closest to $\theta_i(t)$. The bins wrap around in a modular fashion, so that in the example, $\theta_i(t)$ falls in bin 1, rearrangement will happen with its two neighbouring bins: bin 2 and bin 8.

$$k_{temp2} = (0 + 2o_{re}, \max(0,7 - o_{re}),2,2,0,0,0, \max(0,1 - o_{re}))$$

In the example where $o_{re} = 1$, after rearrangement

$$k_{temp2} = (2,6,2,2,0,0,0,0)$$

Fibre deposition

Finally,it was shown in Figure 5 that matrix fibre orientation is highly correlated with the fibroblasts depositing the fibres. Therefore, a fibroblast will deposit fibres in the bin k which is closest to $\theta_i(t)$ if $0 \leq \theta < \pi$, or $\theta_i(t) - \pi$ if $\pi \leq \theta < 2\pi$, since fibres are apolar. In the example, $\theta_i(t) = \frac{\pi}{1000}$ is closest to bin 1 which has orientation 0. Therefore, after fibre deposition with $o_{dep} = 1$

$$k = (3,6,2,2,0,0,0,0)$$

These modifications all occur within a single time step.

3.2.2 Relating cells to matrix

As described in section 3.2.1, the ECM consists of a fine grid, with the distance between grid points being at least as small as a cell's head. At time t, a cell i with orientation $\theta_i(t)$ is associated with and will interact with whichever grid point k is closest to the centre of its head (Figure 38). In this way, a cell is always associated with exactly one grid point.

The organisational and matrix guidance components of the model are implemented algorithmically and best illustrated through an example.

3.2.3 Matrix organising fibroblasts

In this adaptation of the model, a cell's orientation will be calculated as a weighted function of not only individual migratory noise and influence of local neighbours but also of the matrix fibres. The initialisation of the model remains the same as in the previous chapter, together with computation of individual migratory noise and cell-cell guidance. Matrix feedback is implemented as follows:

If the head of a cell i is above a fibre k with orientation ϕ_k, then the X and Y-components of the effect of matrix guidance of k on i at time t are

$$X_{i,m}(t) = \cos\left(\tilde{\phi}_k(t)\right), \qquad Y_{i,m}(t) = \sin\left(\tilde{\phi}_k(t)\right), \tag{26}$$

where

$$\tilde{\phi}_k(t) = \begin{cases} \phi_k(t), & \text{if } | \phi_k(t) - \theta_i(t) | \ (\text{mod } \pi) < \frac{\pi}{2}, \\ \phi_k(t) + \frac{\pi}{2}, & \text{otherwise,} \end{cases} \tag{27}$$

representing the ability of fibroblasts to move along fibres in a nematic manner.

The X and Y-components of cell i at time t as originally described in 2.5.2 in the previous chapter are then redefined to include the influence of matrix feedback so that

$$X_i(t) = \frac{1}{w_p + w_c + w_m}\left(w_p X_{i,p}(t) + w_c X_{i,c}(t) + w_m X_{i,m}(t)\right), \qquad Y_i(t) = \frac{1}{w_p + w_c + w_m}\left(w_p Y_{i,p}(t) + w_c Y_{i,c}(t) + w_m Y_{i,m}(t)\right), \tag{28}$$

is then computed as

$$\theta_i(t + \Delta t) = \tan^{-1}\left(\frac{Y_i}{X_i}\right) \tag{29}$$

which must then be adjusted for quadrant of the arctan function so that

$$\theta_i(t+\Delta t) = \begin{cases} \theta_i(t+\Delta t) & \text{if } X_i \geq 0, Y_i \geq 0, \\ \theta_i(t+\Delta t)+\pi & \text{if } X_i < 0, \\ \theta_i(t+\Delta t)+2\pi & \text{if } X_i \geq 0, Y_i < 0. \end{cases} \quad \textbf{(30)}$$

Finally, cell position is updated so that

$$\begin{bmatrix} x_i(t+\Delta t) \\ y_i(t+\Delta t) \end{bmatrix} = \begin{bmatrix} x_i(t) + s_i \Delta t \cos\big(\theta_i(t+\Delta t)\big)\, v_e \\ y_i(t) + s_i \Delta t \sin\big(\theta_i(t+\Delta t)\big)\, v_e \end{bmatrix} \quad \textbf{(31)}$$

where v_e is a proxy for volume exclusion defined by

$$v_e = \begin{cases} 0.25 & \text{if the head node of a cell is overlapping with any other node of another cell,} \\ 1 & \text{otherwise} \end{cases}$$

The distance between cells, which is considered as an overlap is set by the user, but for the rest of this work was set as the inner 75% of a cell's area.

The order of events in the simulations at each time step is given as:

1. Compute new cell orientations
2. Compute and update changes to matrix grid points
3. Compute and update new cell positions.

A schematic indicating the main model components is given in Figure 40: Full two-layer model schematic.

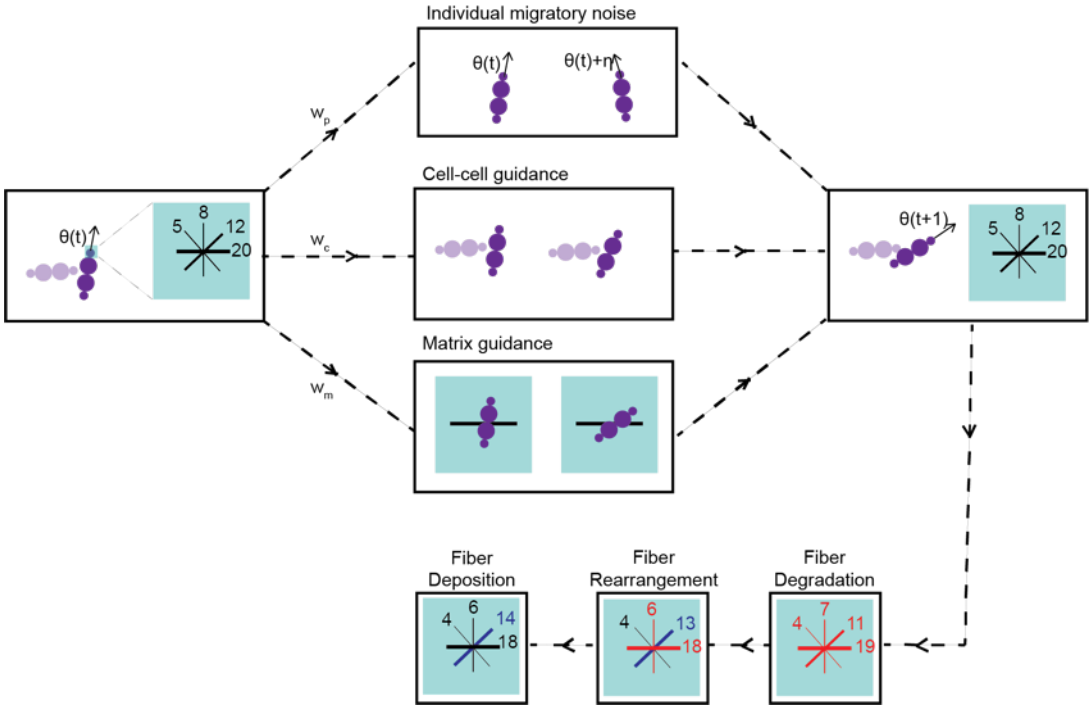


Figure 40: Full two-layer model schematic

Cells flock with other cells they come into contact with and the fibres in the matrix grid point underneath the head of the cell. At every time step, for each cell in turn the model computes its change in orientation due to flocking with other cells, its change in orientation due to flocking with the matrix below and its reorganisation of the matrix in the grid point below *via* degradation, rearrangement and deposition of fibres. In this schematic, there are four bins per grid point for fibres to be deposited in.

A complete list of free parameters of the two-layer model is given in Table 5. Unless otherwise simulations take parameters with the default values given in the "Typical values" column. The primary parameters that will be varied are individual migratory noise (η), cell-cell collision guidance w_c and cell-matrix guidance w_m.

Table 5: List of parameters II

Parameter	Meaning	Typical value
ρ	Number of cells in simulation	800 (corresponding to ~40% confluence)
d	Number of time steps required for population to double	∞ (ie there is no proliferation)
μ	Mean of Gaussian distribution for cell speed	$32\mu m/h$

σ	Standard deviation of Gaussian distribution for cell speed	$4\mu m/h$
a	Number of body nodes (excluding the head and tail nodes), determining aspect ratio	2
r	Radius of cell body nodes (cell area is then $ar^2 + \frac{r^2}{2}$)	$13.5\mu m$ (cell area is then $1410\mu m^2$
η	Individual migratory noise. At each time step $\eta_i(t) \sim N(0, \eta)$	varied
w_c	Weighting of flocking mechanism (collision guidance)	0.03
w_m	Weighting of matrix guidance	Varied
o_{dep}	Number of fibres deposited by a fibroblast in a single time step	1
o_{deg}	Number of fibres removed from each bin in the grid point corresponding to a fibroblast in a single time step	0
o_{re}	Number of fibres moved from a neighbouring bin in the grid point corresponding to a fibroblast in a single time step	0

3.2.4 Model implementation

The C++ code describing the model of Chapter 2 was augmented to incorporate ECM as described above. Figure 41 shows the structure of this extended code. Additional parameters are included for fibre deposition, rearrangement and degradation rates, matrix grid point size and bins per matrix grid point. A new class called Matrix grid point was included in the model, and each grid point was endowed with positional coordinates, bins with an attached fibre density and the dominant density and orientation of the grid point. All of this information could then be generated as an output in the matrix grid map file.

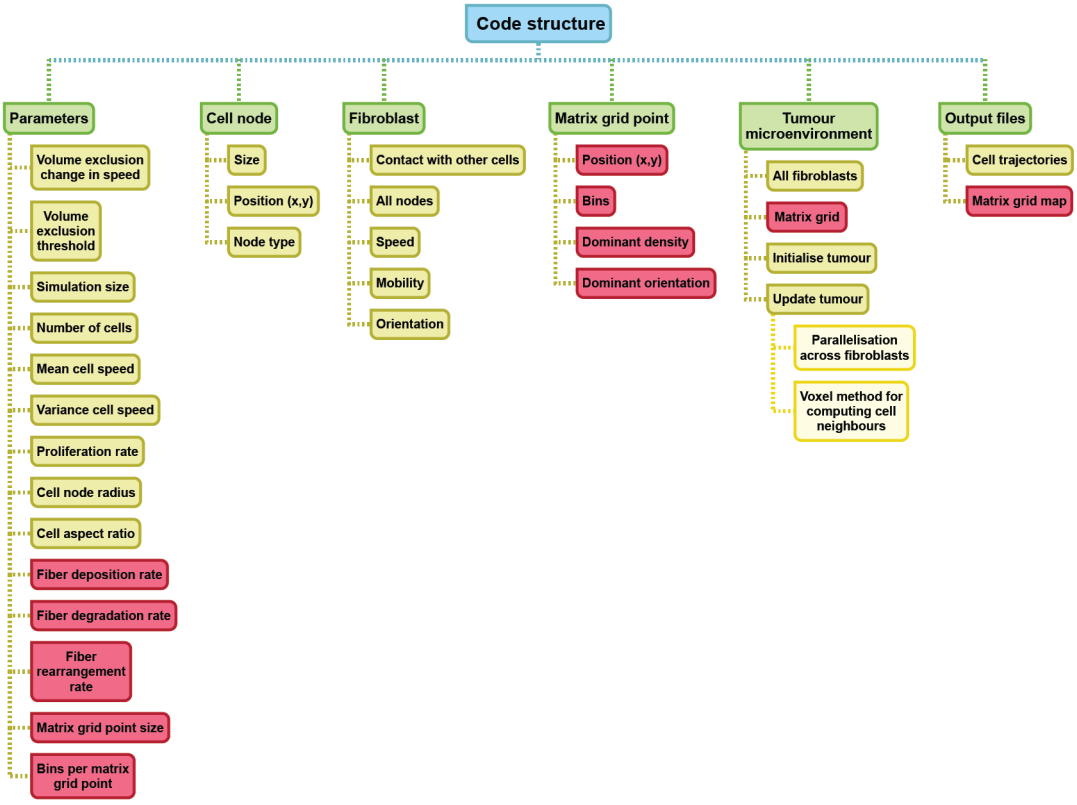


Figure 41: Code structure II

Diagram shows the hierarchical structure of the C++ code for the two-layered model. Modules shown in red are additions to the code described in Chapter 2.

The main update function of the model was subsequently changed from “Update fibroblasts” to “Update TME” to account for updating both fibroblasts and matrix. Algorithmically, it was important for cells and matrix to influence each other. For example, if cells altered all matrix fibres to their orientation and then were guided by those fibres then the influence of matrix on cells would be null. Therefore, the new fibroblast orientation was computed first and then the fibres would be rearranged (Figure 42). A more sophisticated algorithm could have taken a leap-frog method, splitting each time step in two so that the first half time step could be based on the cell’s previous orientation and the second half time step could use the cell’s new orientation. Ultimately, the effect of these choices is minimal provided the time step used is small so that only small changes in cell and ECM orientation occur at each time step.

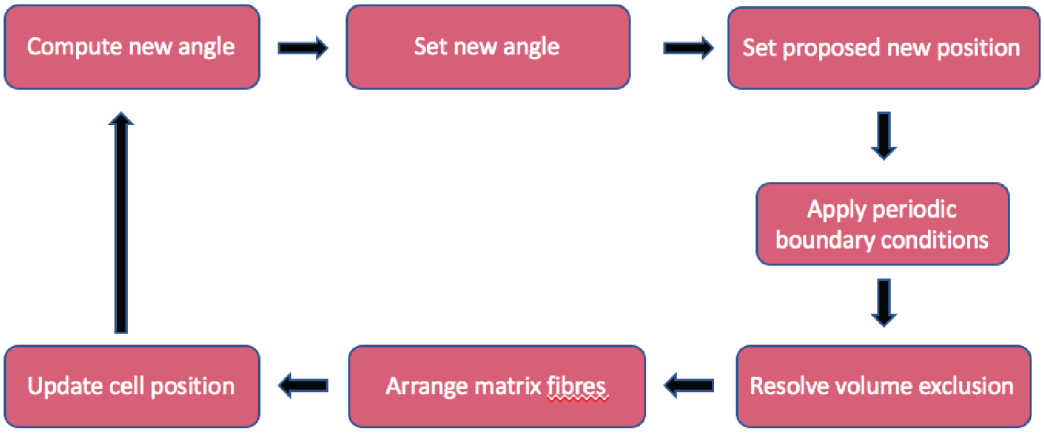


Figure 42: Flow diagram model II

Flow diagram shows sequence of functions carried out by the code at each time step in order to work out the orientation of the fibroblasts and the configuration of the ECM.

Using the output matrix grid map file, it was possible to visually represent the matrix in OpenGL. For each grid point, a straight line indicating the orientation of the grid point's dominant bin was drawn, with the darkness of the line representing the density of fibres in that dominant bin up to a cut off of 25 fibres, represented by a black line.

In the simulation set up, the grid points describing the matrix are initialized at the beginning of the simulation. The coarseness of the overall matrix grid is determined by the number of grid points from which it is constructed. By default, the number of grid points is computed as

$$\#grid\ points = \left(\frac{simulation\ box\ length}{diameter\ of\ cell\ head}\right)^2 \quad \textbf{(32)}$$

which equates to a total of $\left(\frac{1024}{8}\right)^2 = 16384$ squares.

The matrix images produced *in silico* show the density of the most recently chosen bin at each grid point.

3.2.5 Adapting metrics to in silico matrix

Given the way in which *in silico* matrix was recorded and visualized showing each grid point's dominant bin it was necessary to adapt the matrix alignment metrics used for the *in vivo* images to deal with the alternative *in silico* representation.

Typically, simulations were run with $128 \times 128 = 16384$ grid points. The positions and dominant orientations of the 800 densest grid points was recorded. The alignment of these points was then computed using the same method as was used for cell bodies as described in 2.2.1. This represented considering the alignment of the densest 5% of fibres. This value was chosen to avoid the confounding effects of considering grid points with very few fibres in them. Alternatively, an algorithm where each grid point was weighted according to its density could have been employed, but with much higher computational costs.

The metrics of high-density matrix (HDM), curvature (Curv) and fractal dimension (Frac) were all measured *via* the same methods as used for the *in vivo* images as described in the previous chapter.

3.3 Parameter fitting and robustness

3.3.1 Fitting matrix feedback

Matrix feedback (w_m) was fitted in a similar way to cell-cell collision guidance as described in 3.1.3. In the experiment described in Figure 36, control fibroblasts were seeded in sub-confluent conditions on glass. A second set of fibroblasts at the same level of confluence were seeded on pre-existing thick aligned matrix. These fibroblasts displayed increased migrational persistence, confirming the effect of matrix feedback on cell migration. In order to quantify the effect of matrix feedback, the persistence of these two groups of fibroblasts was computed for windows of one and two hours. Cells tracked for at least one hour were recorded and a spline tracing the cell's trajectory was produced using the loess package in R ($\alpha = 0.5$) as described in section 2.6.1. This was in order to smooth the intracellular movement,

which resulted in many small fluctuations in the trajectories. For each cell, the median persistence for each cell was computed as described above. Simulations of single cells with varying values of individual migratory noise (η) were run and the persistence computed for one and two-hour windows. The experimental results were then matched with simulations using a least squares approach to select the most likely value of η given a cell's persistence. The distribution of noise in the simulations that matched the cells moving on glass was $\eta = 0.13$. Simulations were then run for the same level of noise and incremental values of w_m between 0 and 1. The persistence of these simulations was then compared with the persistence of the cells moving on thick aligned matrix. Using a least squares approach, the most likely value of matrix feedback on a thick matrix is approximately $w_m = 0.27$.

3.3.2 Parameter robustness

The key parameters of interest in the model are individual migratory noise, cell-cell collision guidance and matrix feedback. However, there are many other parameters in the model and it is useful to consider to what extent the model is sensitive to these other parameters of secondary interest.

3.3.3 Cell shape

In the simulations, cells are endowed with a fixed diamond shape consisting of a smaller bead, two larger beads, followed by a final smaller bead. As shown in other works, cell shape is in fact variable and dynamic over time (Keren *et al.*, 2008; Tweedy *et al.*, 2013; Yin *et al.*, 2013). As described in Yin et al., cell shape could be characterised as belonging to one of five discrete categories: rounded, elongated, large cells with smooth edges, teardrop-shaped cells and very large flat cells with ruffled edges. Further, Figure 39 appears to show fibroblasts that have a more teardrop like morphology. Therefore, to probe the effect of cell shape on emergent matrix pattern, simulations were run with three of these cell shapes: rounded, teardrop and elongated. These cell shapes were constructed by assembling different sized beads and all cells were scaled to be approximately the same size across shapes. The smoothness of cell edges was not considered to be

of importance at the mesoscale of the model, thus merging three of the categories identified in Yin et al. into a single “rounded” category.

With these three cell shapes simulations were run, exploring how long-range alignment (LRA) changes through parameter space, varying individual migratory noise (η), cell-cell guidance (w_c) and matrix feedback (w_m). Results are shown in Figure 43. In each heatmap, one of the three parameters is fixed at zero and the other two parameters are adjusted incrementally. Comparing the heatmaps row-wise shows that there are only minor differences in alignment depending on cell shape. The rounded cells produce slightly more aligned patterns, which could be explained by the increased surface area of the head bead (which in this case is the entire cell), corresponding to a wider neighbourhood in the traditional Vicsek model. However, for these simulations, cell speed was the same and it is in fact well documented that there is a direct coupling between cell speed and cell persistence (Maiuri et al. 2015, Cell 161, 374–386). Simulations were run with rounded CAFs which move very slowly and have high migratory persistence (Figure 44). If cells move too slowly or are too small, they do not come into contact frequently enough to generate alignment. Therefore, whilst the rounded cells in the simulations are marginally better at producing alignment, it is unlikely in biological systems that cells with this morphology could move quickly or with persistent migration.

The analyses in Figure 43 confirm that matrix patterning is largely robust to different cell shapes at the mesoscale. For this reason, the chosen diamond cell shape in the model is a reasonable biological assumption, describing a cell with elongated morphology with aspect ratio 1:3, which is the average aspect ratio taken from experimental data (Figure 10). In addition, simulations varying aspect ratio were performed and confirm that altering aspect ratio does not greatly alter emergent patterning (Figure 45). Only a modest decrease in HDM and small increase in fractal dimension are observed for the extreme case of fibroblasts being entirely rounded (aspect ratio = 1).

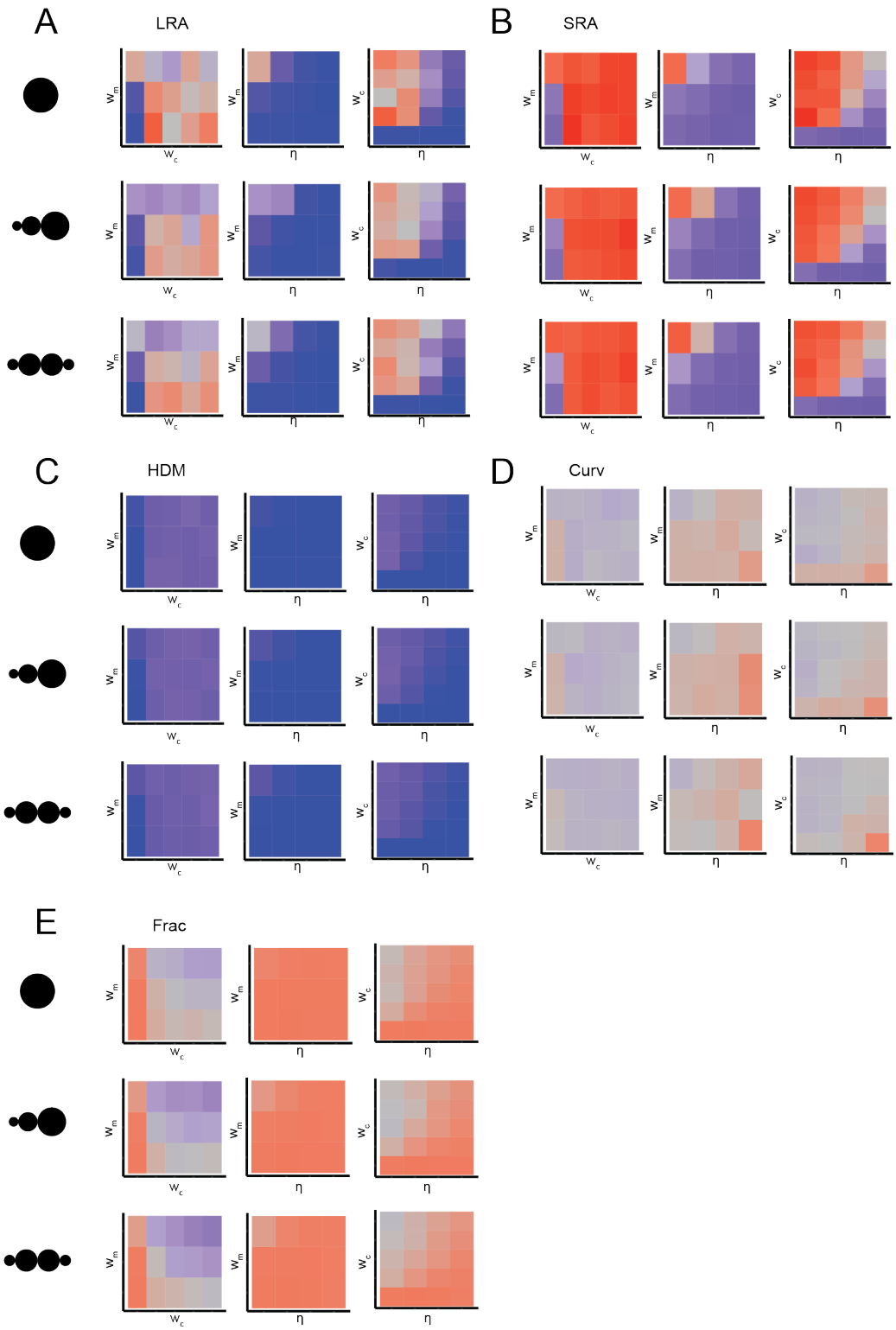


Figure 43: Exploring the effect of cell shape

(A) Heatmaps showing long-range alignment (LRA, $0 - 400\mu m$) for simulations with CAFs with an elongated, teardrop and rounded morphology (top, middle and bottom rows respectively). Schematics of these cell shapes are shown on the left. In the first column of heatmaps, matrix feedback is fixed at zero ($w_m = 0$) whilst noise (η) and cell-cell guidance (w_c) are varied incrementally. In the second

column, ($w_c = 0$) whilst η and w_m are varied and in the third column, $\eta = 0$ whilst w_c and w_m are varied. Comparing the heatmaps row-wise shows that a different cell shape causes little difference in LRA. N=5 simulations per point in parameter space. Simulations are of 500 cells. For all simulations, matrix organisation parameters are set to the typical values given in Table 5: $o_{dep} = 1, o_{deg} = 0, o_{re} = 0$. Parallel analysis is done for short-range alignment (SRA), high-density matrix (HDM), curvature (Curv) and fractal dimension (Frac) in figures B, C, D and E respectively.

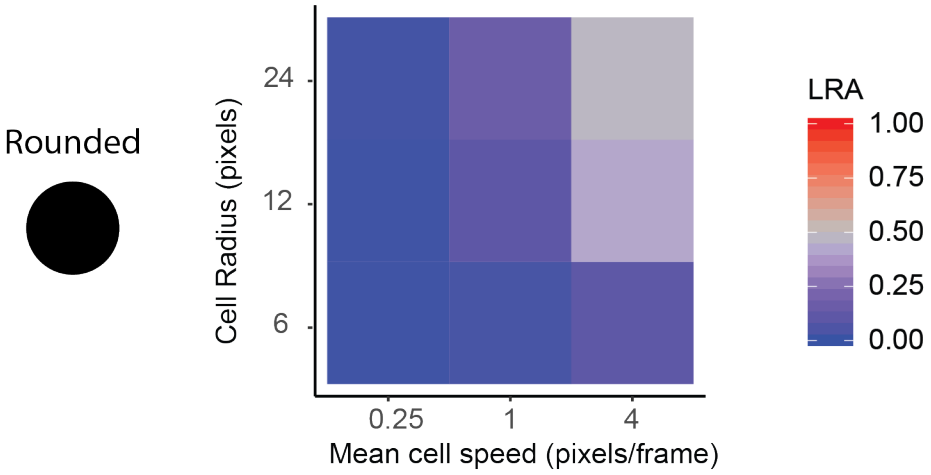


Figure 44: Varying cell speed and size

Heatmap showing long-range alignment (LRA, $0 - 400\mu m$) for simulations for varying mean cell speed and cell radius. Cell shape is rounded, as shown schematically on the left. N=5 simulations per point in parameter space. Parameters are fixed as follows: $\eta = 0.02, w_c = 0.05, w_m = 0$. All other parameters taken the default values given in Table 5.

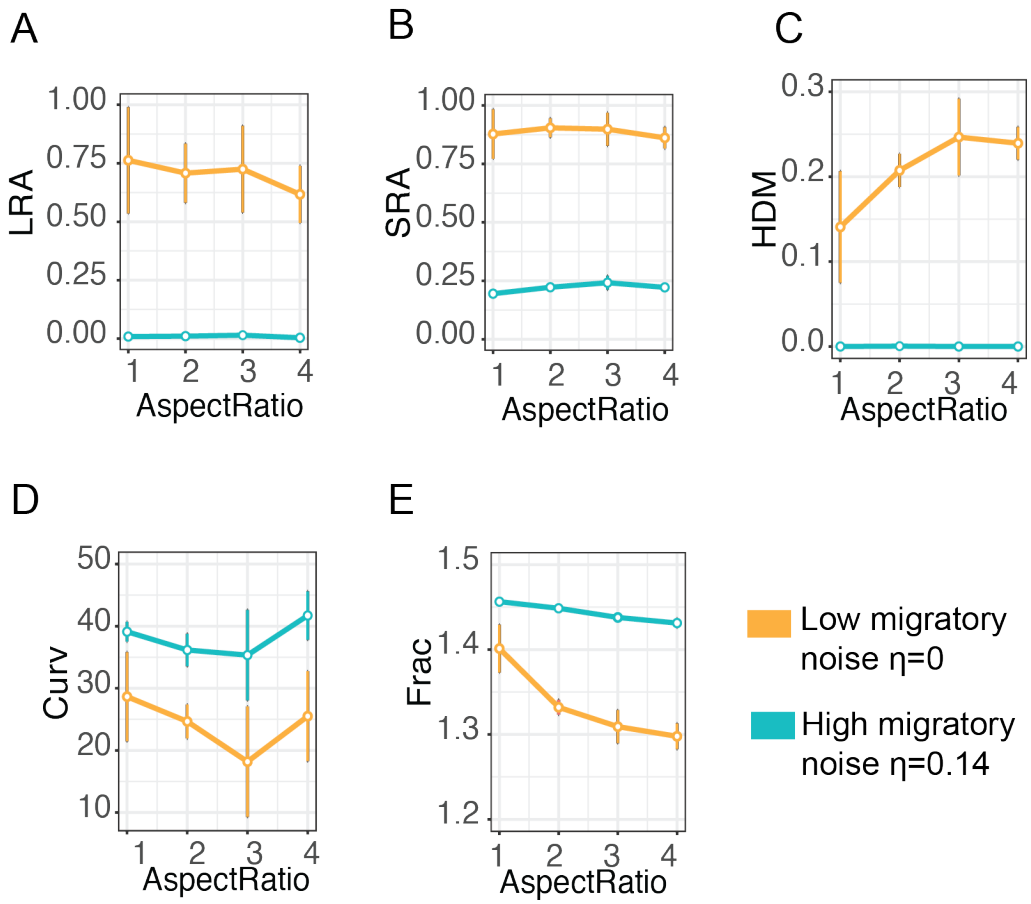


Figure 45: Varying aspect ratio

The effect of increasing cell aspect ratio on matrix organisation for cells with low individual migratory noise ($\eta = 0$, orange) and high individual migratory noise ($\eta = 0.14$, blue). Parameters follow typical values indicated in Table 5 and $w_c = 0.03, w_m = 0$. N=5 simulations per point in parameter space. Error bars show 95% confidence intervals. Simulations run with 800 cells.

3.3.4 Cell speed

One assumption of the model is that each cell moves at a fixed speed, determined at its initiation from a Gaussian distribution with mean and standard deviation derived from experimental data. The parameter of cell speed can be directly measured from experimentation and translated to the mathematical model. However, it is likely that cells move at a different speed *in vivo*. Altering cell speed can cause variations in matrix pattern. If cell speed is very high, then fibroblasts will jump across multiple matrix grid points in a single time step, creating a dashed line of fibres. Conversely, if cell speed is very slow, then a cell may deposit fibres in the same grid point over several time points, leading to thicker matrix. In the case of

cells that rearrange matrix fibres, if the cells are much slower moving then they are likely to rearrange more of the fibres that they pass over. An interesting discussion of the effect of particle speed on emergent behaviour can be found in previous literature (Dallon, Sherratt and Maini, 1999).

3.3.5 Matrix grid layer parameters

In the simulations, the matrix grid layer comprises of a user-defined number of grid points. The number of bins within each grid point is also determined by the user. For simulations shown in Figure 51 there were 128x128 grid points, so that each grid point is approximately the area of the cell head bead. Increasing the number of grid points increases the precision of the matrix organisation and deposition (Figure 46). This comes at a cost of increased computational time. Further, as the number of grid points is increased, average fibre density per grid point will decrease and time step will have to be reduced to avoid cells jumping over many grid points in one time step.

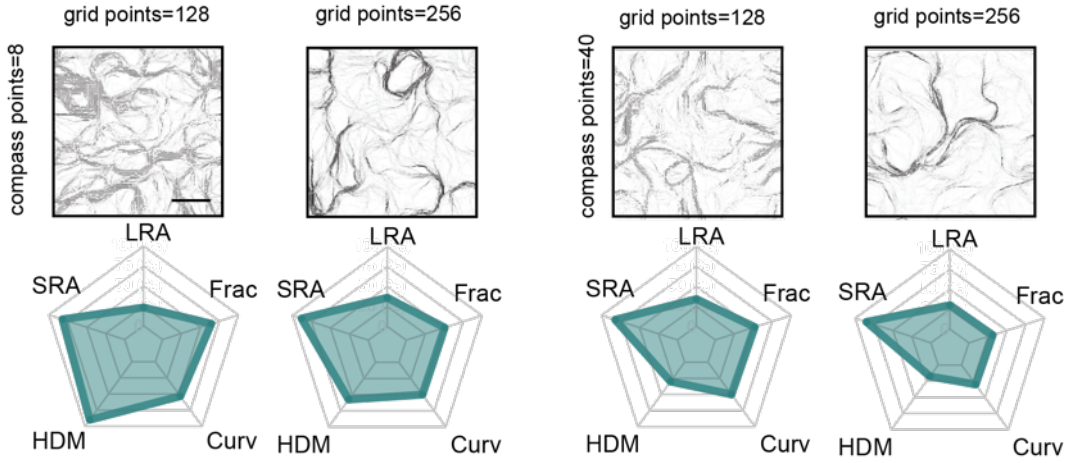


Figure 46: Varying matrix grid parameters

Example stills varying number of matrix grid point and the number of bins per grid point with corresponding starplots below. Scale bar represents $100\mu m. \eta = 0, w_c = 0, w_m = 0.1$.

Table 6: Metric values corresponding to Figure 46

	LRA	SRA	HDM	Curv	Frac
noBins = 8, grid points = 128	0.23	0.82	0.36	26	1.307
noBins = 8, grid points = 256	0.36	0.88	0.23	25	1.253
noBins = 40, grid points = 128	0.34	0.82	0.12	25	1.258
noBins = 40, grid points = 256	0.3	0.85	0.07	19	1.176

Analysis indicated that there was no discernible difference between simulations with 8 or 40 bins per grid point (Figure 46). As with the number of grid points, as the number of bins increases, average fibre density per bin will decrease. At an extreme, if there were only two bins per grid point, the matrix would be limited to a basket-weave pattern as fibres would only be oriented north-south or east-west.

Importantly, the possibility that curvature was caused by the number of grid points comprising the matrix or the number of bins at each grid point was excluded, by running a subset of simulations with finer grid points and more bins. It would however be of interest to take this research forward in conjunction with experimentation studying long-range contractility, in order to try to ascertain what a reasonable grid point coarseness would be. If cells are capable of rearranging fibres over long distances, a coarser matrix grid would be appropriate.

3.3.6 A general discussion on parameter sensitivity

Even with the relatively tractable model introduced in this work, it is very easy to introduce many parameters. In addition to the three parameters of interest (individual migratory noise, cell-cell collision guidance and matrix feedback), there are many other secondary parameters (cell speed, size, aspect ratio, matrix grid size, number of bins etc.). The effect of secondary parameters must be considered. Where possible experimental values have been directly used in simulations. Exploratory simulations have shown that the parameters for which it was not possible to derive experimental data have little effect on matrix pattern, for example matrix grid size. In any case, care has been taken to try to choose feasible parameter values.

3.4 *In silico* generation of diverse matrix patterns

3.4.1 Individual migratory noise and collision guidance insufficient to generate diverse patterns

Initially the model was used to generate matrix patterns by varying cells' individual migratory noise (η) and collision guidance (w_c) as done in Chapter 2, looking at organisation of the fibroblasts themselves. The fibroblasts deposited matrix fibres onto the second layer of the model but matrix feedback (w_m) was set to zero, indicated the cells moving entirely independently of the matrix fibres. Unsurprisingly, the matrix patterns that can be produced by these competing effects range from isotropic to highly aligned. The top row of Figure 47A indicates that only isotropic matrix is generated in the absence of collision guidance, no matter the level of individual migratory noise of cells. Introducing of collision guidance can lead to the generation of aligned matrix with spatially uniform distribution of the ECM as shown in Figure 47A. Crucially however, the patterns generated by individual migratory noise and collision guidance alone do not include many of the diverse patterns observed *in vivo*. The range of patterning is limited to alignment and disorder. In particular, "swirl-like" patterns across a mesoscale, as seen in the dermis and liver, with high curvature and short-range alignment could not be recapitulated by the model. Having established that fibroblasts were indeed guided by ECM (Figure 36) it was hypothesized that this could have a significant effect on emergent matrix patterns.

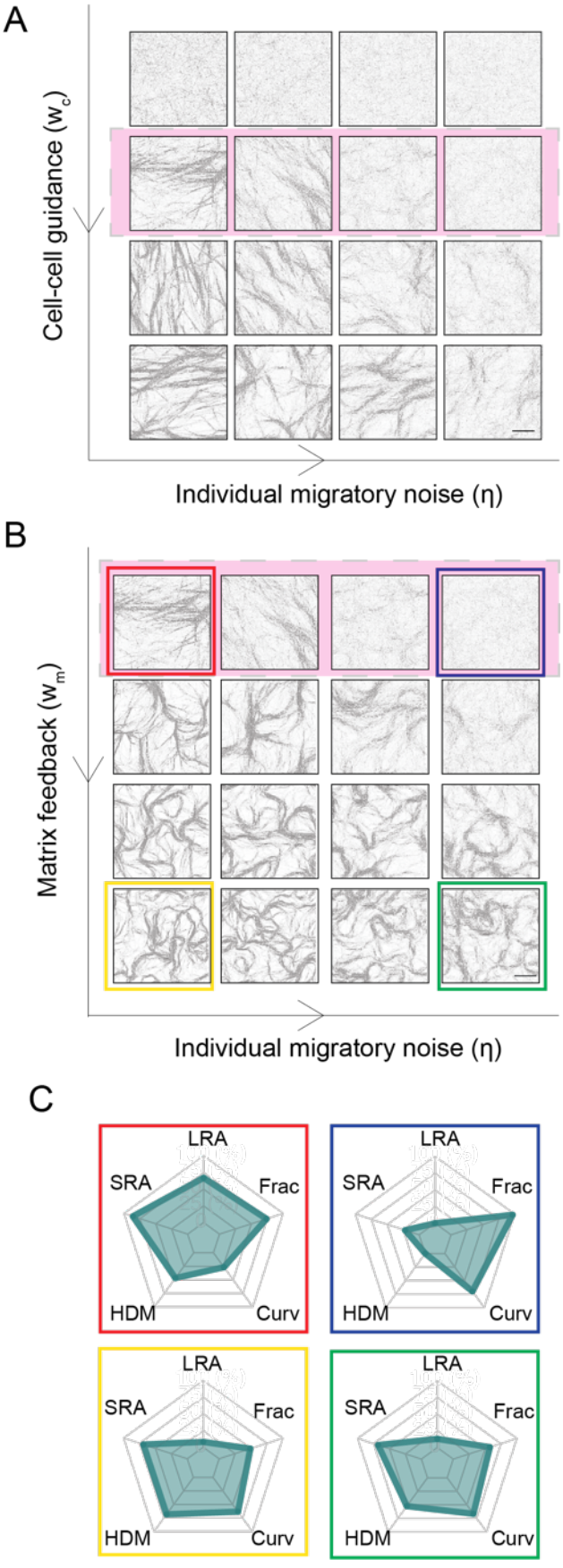

A
Cell-cell guidance (w_c)
Individual migratory noise (η)
B
Matrix feedback (w_m)
Individual migratory noise (η)
C
LRA
SRA
Frac
HDM
Curv

Figure 47: Matrix feedback increases diversity

Comparison of emergent *in silico* matrix patterns from cell-cell interactions with cell-matrix interactions. (A) Matrix patterns produced from cell-cell interactions alone with varying noise and cell-cell guidance. From left to right noise $\eta = (0, 0.07, 0.14, 0.21)$, from top to bottom $w_c = (0, 0.03, 0.06, 0.1)$. The pink row, showing cell-cell guidance $w_2 = 0.03$, is repeated in part (B). Scale bar represents $100\mu m$. (B) Matrix patterns produced from varying noise and cell-matrix feedback, cell-cell guidance fixed at $w_c = 0.03$. Simulations are of 800 cells over a time-course of seven days. From left to right noise $\eta = (0, 0.07, 0.14, 0.21)$, from top to bottom $w_m = (0, 0.04, 0.12, 0.2)$. (C) Corresponding starplots showing metrics characterising distinct matrix patterns.

3.4.2 Matrix feedback generates diverse matrix patterns

To test this hypothesis, first matrix organisation was kept constant whilst the strength of the term describing how the matrix topology influences the cell orientation (w_m) was varied. Cells would deposit a single fibre per time step with no degradation or rearrangement of fibres. Given the restricted selection of patterns that can be generated by cell-cell guidance alone, this parameter was fixed at $w_c = 0.03$, reflecting a physiologically plausible level of coordination between cells such that cells with zero or low individual migratory noise will align but cells with high individual migratory noise will not, thus producing an isotropic matrix. This value of collision guidance is similar to the value derived from experiments of the aligning CAFs. Fixing these parameters enabled for a simpler exploration of the effect of matrix feedback on resulting ECM patterns. Matrix feedback in the model (w_m) was varied incrementally. Figure 47B shows that varying matrix feedback greatly increased the diversity of matrix patterning.

In the extreme cases of varying individual migratory noise and matrix feedback between being very low or very high demonstrates the diversity of patterns that can be achieved. In the case in which cells have low individual migratory noise and zero matrix feedback, the resulting emergent matrix is aligned characterised by high LRA, high SRA and low Curv (Figure 47B and C red box). Cells with similarly low levels of individual migrator noise but high matrix feedback produce matric with high SRA but low LRA, medium curvature and a high HDM (Figure 47B and C, yellow box).

As documented previously, cells with high individual migratory noise and zero matrix feedback produce a diffuse isotropic matrix with high Frac and Curv and low LRA and SRA (Figure 47B and C, blue box). Cells with similarly high level of individual migratory noise but high matrix feedback generated matrix with low LRA, quite high SRA and a mid-range percentage of HDM and Curv (Figure 47B and C, green box).

	LRA	SRA	HDM	Curv	Frac
Figure 47A, red	0.67	0.86	0.19	18	1.345
Figure 47A, blue	0	0.22	0	31	1.437
Figure 47A, yellow	0.08	0.69	0.27	29	1.246
Figure 47A, green	0.13	0.68	0.21	30	1.28

Figure 48 shows stills from example simulations demonstrating the coevolution of pattern of the fibroblasts and matrix. The correlation between cell body organisation and underlying ECM can clearly be observed.

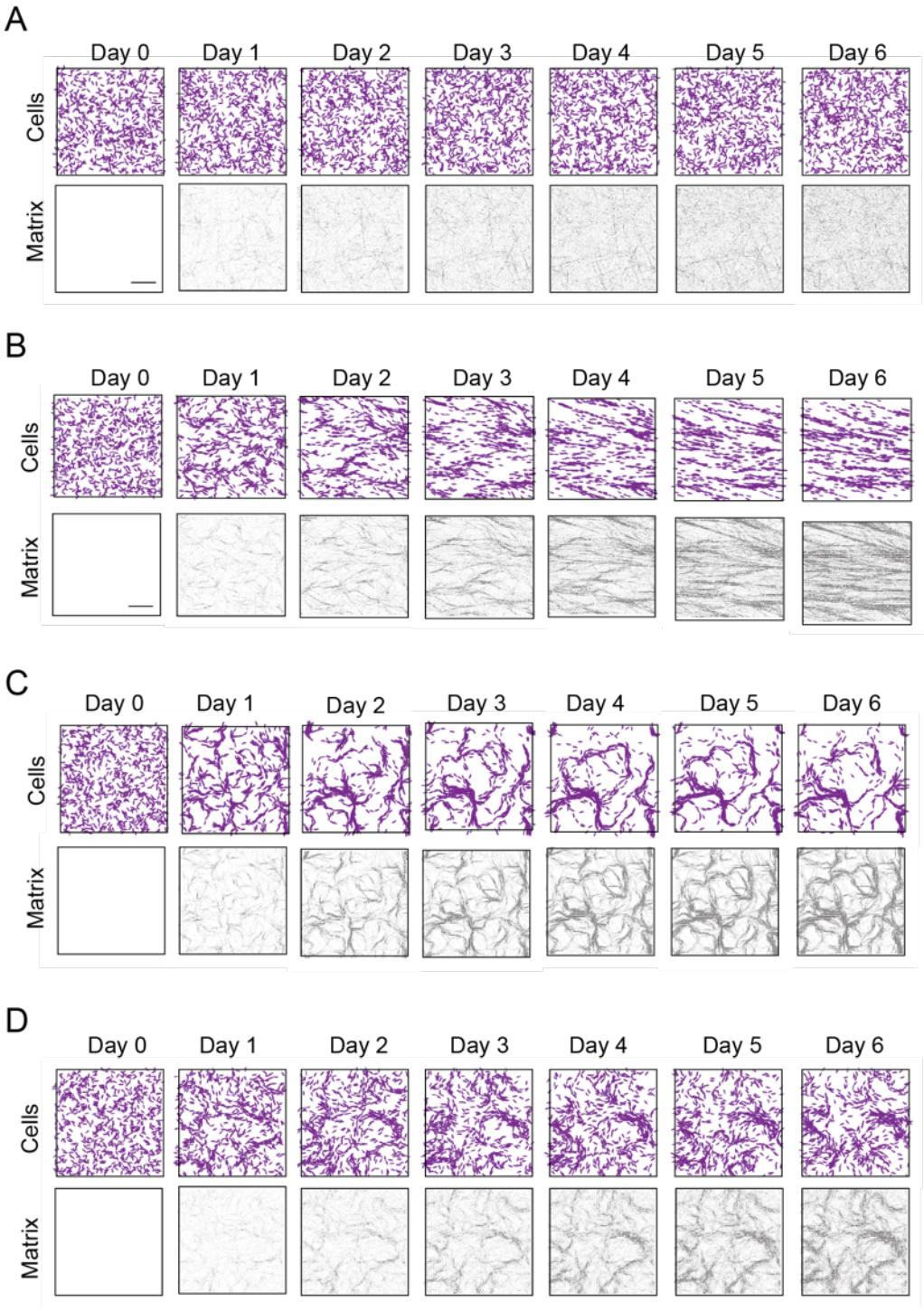


Figure 48: Simulation stills of emergent patterns

Matrix patterns emerging over time. Images from simulations showing fibroblasts (top) and corresponding matrix (bottom) over six days. (A) Isotropic matrix generated with parameters $\eta = 0, w_c = 0, w_m = 0$. (B) Aligned matrix generated with parameters $\eta = 0, w_c = 0.03, w_m = 0$. (C) Swirly matrix generated with parameters are set at $\eta = 0, w_c = 0.03, w_m = 0.2$. (D) Diffuse swirly matrix generated by $\eta = 0.14, w_c = 0, w_m = 0$. For all simulations deposition rate=1, degradation rate=0, rearrangement rate=0. Scale bar represents $100\mu m$.

To investigate further the effect of matrix feedback simulations were run varying individual migratory noise, cell-cell guidance and matrix feedback together. Pairwise analysis of the metrics revealed that the addition of matrix feedback could produce patterns in new areas of metric-space (Figure 49). In particular, matrix could now be produced with low long-range alignment but high short-range alignment, high percentage of high-density matrix (~30%) and a wide range of curvatures. This describes the spectrum of "swirl-like" patterns created by matrix feedback.

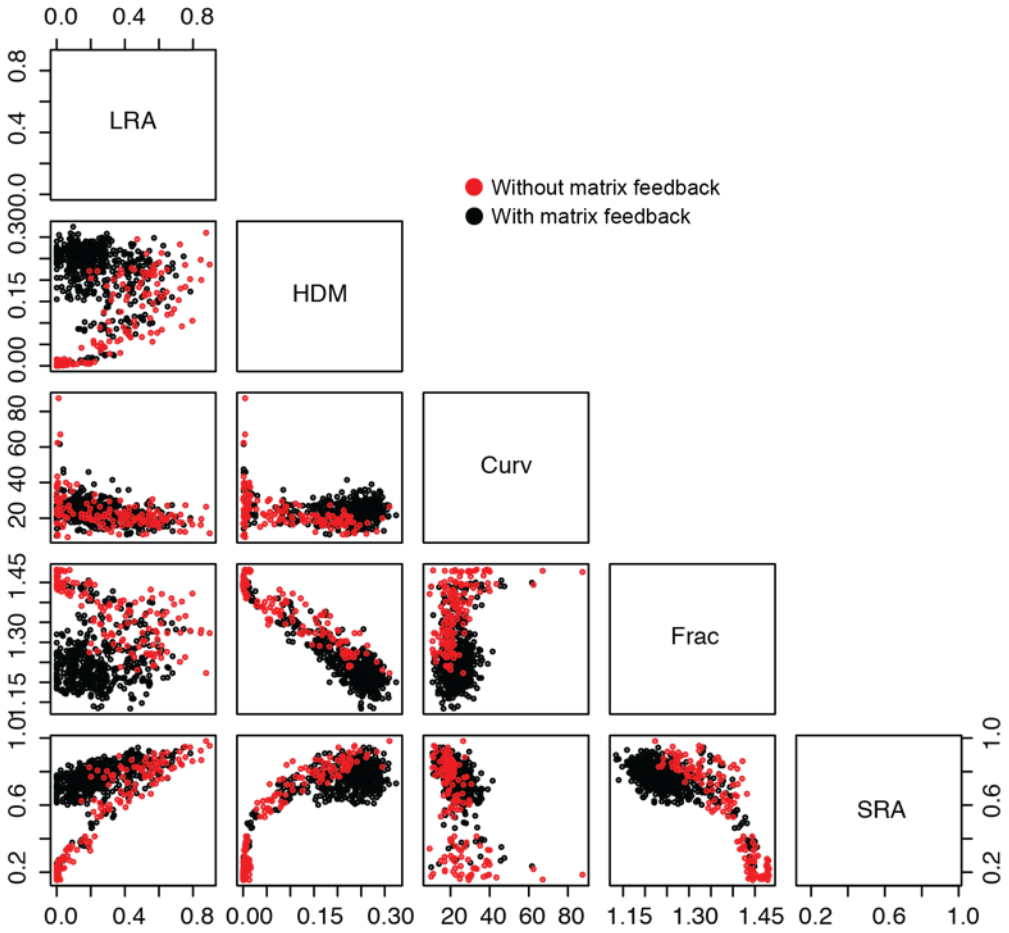


Figure 49: Pairwise analysis of metric space

Pair-wise analysis comparing metric-space covered by cells without matrix feedback (red) and with matrix feedback (black) showing the differences between patterns. N=10 simulations per point in parameter space. X and Y-axes for each metric shown. Matrix patterns produced from varying noise, cell-cell collision guidance and cell-matrix feedback ($\eta = (0,0.05,0.1,0.14), w_c = (0,0.03,0.06,0.1), w_m = (0,0.04,0.12,0.2)$). Simulations are of 800 cells over a time-course of seven days.

These new patterns that are produced with the addition of matrix feedback arise due to the increased corralling of the cells as matrix guidance causes them to follow and reinforce tracks of matrix fibres that were laid down at early time points. Visually, the result is pronounced swirl-like patterning. This corralling of cells causes cells with low individual migratory noise to produce matrix patterns with higher curvature. Corralling also causes a decrease in fractal dimension of matrix patterns produced by cells with any level of individual migratory noise. The matrix acts as a memory component to the flocking model, causing cells to be influenced by matrix that had been deposited early on. This can cause swirl-like patterns later on as fibroblasts reinforce existing matrix.

Consequently, for cells with low individual migratory noise, matrix feedback actually antagonises the matrix alignment that would result from the collision guidance term alone (Figure 50A and B) as cells are influenced by non-aligned matrix deposited early on. Interestingly, matrix feedback could act as a secondary mechanism for generating short-range alignment of cells with high individual migratory noise (Figure 50, blue line). Further, Figure 50 shows that starting with two cell phenotypes (with high (blue line) and low (yellow line) individual migratory persistence respectively), as matrix feedback increases, the patterning of the two cell-types converges. This suggests that matrix feedback can play a dominant role in determining emergent matrix patterning.

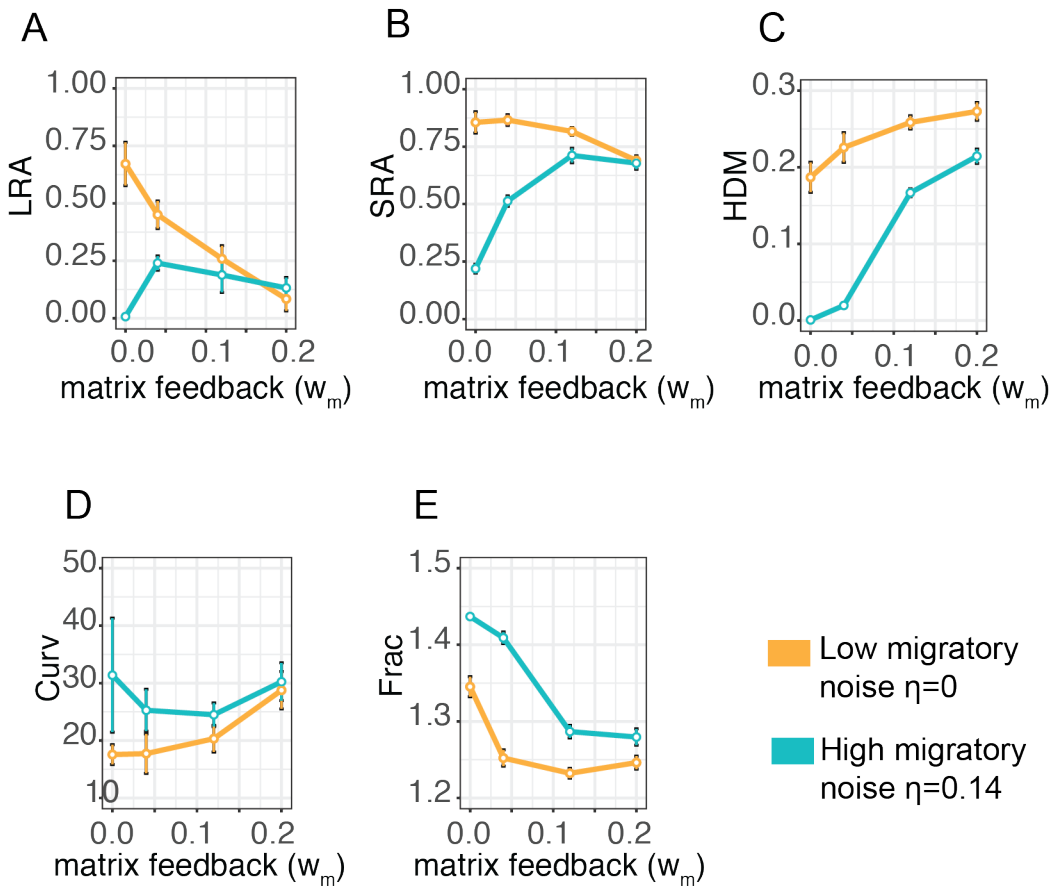


Figure 50: Line graphs showing effect of matrix feedback

The effect of increasing matrix feedback for cells with low individual migratory noise (η=0, orange) and high individual migratory noise (η=0.14, blue). Error bars show 95% confidence intervals. Simulations run with 800 cells and N=20 simulations per point in parameter space.

3.4.3 Dimensionality reduction

Pairwise analysis as described above helps to pinpoint exactly the ways in which matrix feedback affects the structure of ECM. However, in order to gain further depth of understanding as to exactly how matrix feedback alters patterning, a dimensionality-reduction analysis method was required.

PCA with a covariance matrix was carried out to reduce the dimensionality of the five ECM metrics using the pcaMethods package in R (Stacklies *et al.*, 2007), with the nipals method to account for missing data entries (Figure 51). Missing data arose from faint matrix patterns for example, when the matrix was very diffuse and therefore had no mask with which a measurement of curvature could be obtained.

Reducing the data into two principal components could explain 85% of the variance. The loadings plot shows that the first principal component is related to an increase in short-range alignment, high-density matrix and a decrease in fractal dimension, largely describing the effects of increasing matrix feedback. The second principal component related to an increase in long-range alignment with a small reduction in curvature and high-density matrix, resembling the effects of increasing cell-cell guidance. Importantly, the addition of the matrix feedback term expanded the diversity of matrix outputs visible on the PCA plot (note the expanded region covered by the black dots in Figure 51).

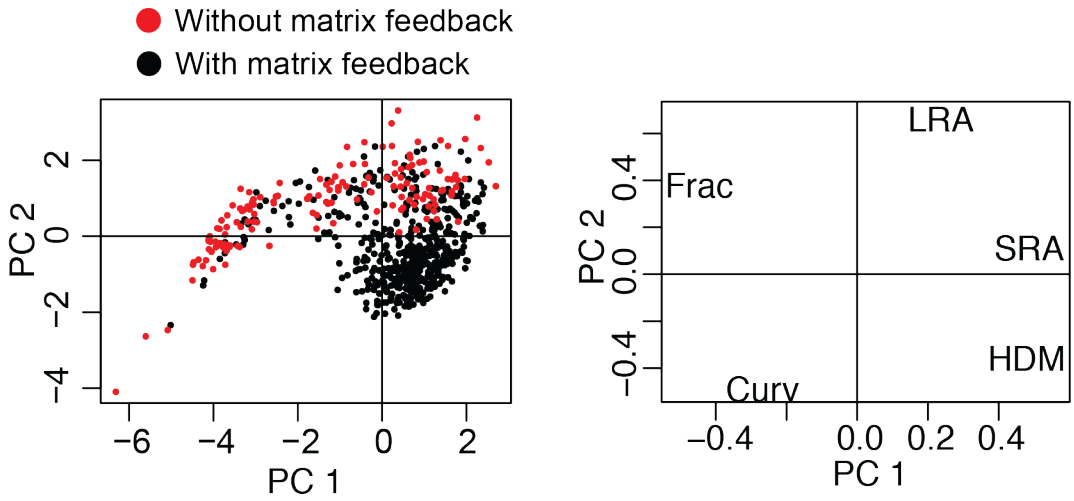


Figure 51: Principal component analysis of metric space

PCA of simulations exploring parameter space, reducing metric-space down into two principal components explains 85% of variance and shows the areas in metric-space that can be reached with matrix feedback. Matrix patterns produced from varying noise, cell-cell collision guidance and cell-matrix feedback ($\eta = (0,0.05,0.1,0.14), w_c = (0,0.03,0.06,0.1), w_m = (0,0.04,0.12,0.2)$).

Similarly, analysis was run in sub-confluent conditions. In the previous analysis, simulations were run with 800 cells, corresponding to approximately 40% confluence. Sub-confluent simulations were run with 100 cells, corresponding to approximately 5% confluence. Again, matrix feedback enhanced diversity of patterns (Figure 52). This level of confluence is perhaps closer to some of the *in vivo* systems in which diverse matrix patterns are observed. As demonstrated in 2.7.5, alignment through cell-cell guidance alone, with zero matrix feedback has a requirement for a threshold confluence. Crucially whilst these sub-confluent

conditions lie below this threshold for alignment, corralling of cells by matrix feedback occurs even at this low confluence, causing emergent patterning differing from complete disorder. Simulations run at sub-confluence have even more pronounced differences in the PCA analysis between matrix patterns produced with matrix feedback as compared to without matrix feedback.

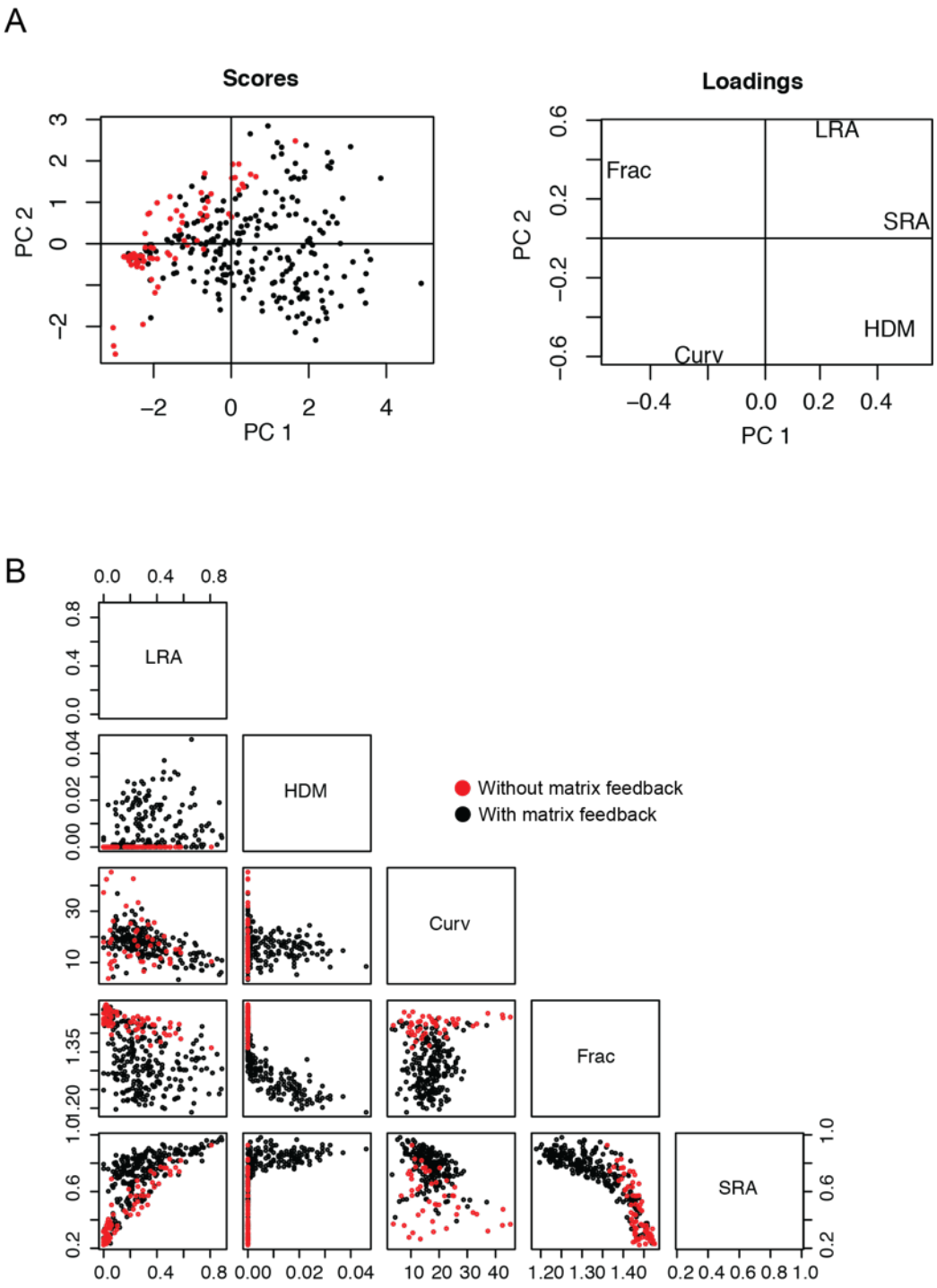


Figure 52: Principal component analysis in sub-confluent conditions

(A) PCA of sub-confluent simulations into two components explains 82% of the variance. (B) Pairwise analysis comparing cells in sub-confluent conditions without matrix feedback (red) against cells with matrix feedback (black) whilst varying cell-cell collision guidance and noise. Simulations are of 50 cells over a time-course of seven days.

3.4.4 Alternative statistical procedures for metric space analysis

PCA with correlation matrix

PCA is typically performed using the covariance matrix of variables. However, results can be misleading if variables have very different scales. The correlation matrix is the standardized version of the covariance matrix. Using the correlation matrix instead accounts for differences in scale, although loses some structure as correlation coefficients are insensitive to variations in data dispersion. PCA using the correlation matrix together with corresponding loadings is shown in Figure 53A. The first principal component relates to an increase in percentage of high-density matrix (HDM) and short-range alignment (SRA), with a simultaneous decrease in fractal dimension (Frac). This change can largely be described by corralling as an effect of matrix feedback, which produces structures swirl-like matrix from cells that would typically form diffuse isotropic matrices without matrix feedback. The second principal component indicates represents an increase in curvature (Curv). In this PCA plot, it is still clear that the addition of matrix feedback enables patterns to reach new areas of metric space with high SRA, HDM and Curv and low Frac.

Convolutional neural network (CNN)

One issue with the analysis methods presented so far is their reliance on user-defined metrics. A degree of subjectivity is introduced when defining the distances over which long and short-range alignment and curvature window are defined. In addition, the metrics in some instances may not be sufficient to describe differences between patterns, as observed in Figure 32, where the starplots characterising the dermis and spleen look similar, despite the *in vivo* images being distinct. Deep learning by contrast provides a metric-free approach to analysing the matrix patterns. Using TensorFlow to create a convolutional neural network (CNN) the algorithm could be trained to classify matrix patterns by whether or not they had been generated with nonzero matrix feedback. Training was carried out with images split into two categories: “With feedback” and “Without feedback”. Based on training with 700 images in each category and 500 training steps the model

achieved 85.2% accuracy in distinguishing between patterns produced with and without matrix feedback.

The different analyses presented here are intended to demonstrate that matrix feedback can generate new patterns that cannot otherwise be obtained through varying just individual migratory noise and cell-cell collision guidance. It also suggests that no single analysis method provides a “silver bullet” for understanding high-dimensional data and that each algorithm may be viewed as providing a new perspective on the data.

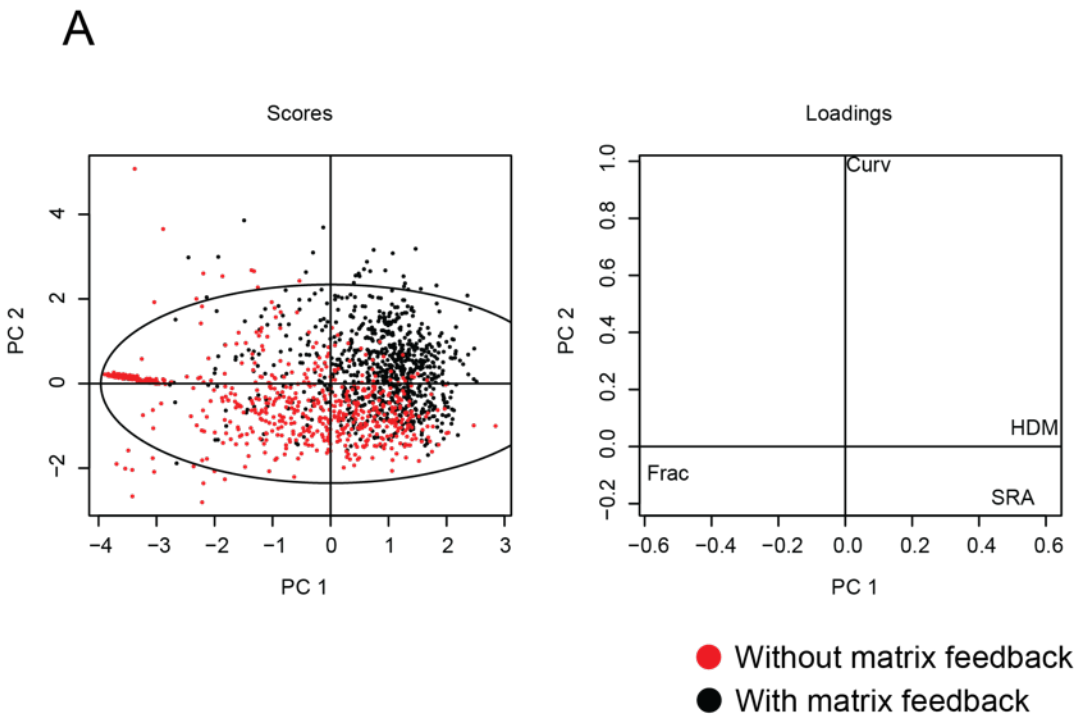


Figure 53: Alternative dimensionality reduction methods

PCA into two principal components using the correlation matrix explains 91% of the variance.

3.5 Exploring fibre organisation

3.5.1 Modelling matrix feedback as a function of fibre density

Figure 28 suggests that for the duration of the four-day FUD experiment (where there is zero matrix guidance) as compared to the control experiment, there was no difference in the persistence or the emergence of alignment. To reflect the idea that matrix feedback has a minimal role at early time-points when little ECM has been deposited, the model was modified to consider matrix feedback as a function of fibre density. Experimental data was used to devise a heuristic argument as to what that function f might be. Then it was possible to redefine the X and Y-components determining fibroblast orientation as written in 2.5.2 as

$$X_i = \frac{1}{w_p + w_c + f(d)w_m}(w_p X_p + w_c X_c + f(d)w_m X_m), \quad \textbf{(33)}$$

$$Y_i = \frac{1}{w_p + w_c + f(d)w_m}(w_p Y_p + w_c Y_c + f(d)w_m Y_m),$$

In the FUD experiment (Figure 28), there was no difference in the persistence or the emergence of alignment of the FUD knockout cells as compared with the control fibroblasts. Therefore, the assumption is made that the control fibroblasts undergo little to no matrix guidance with matrix fibres produced for these first four days. This is defined mathematically such that for days 1-4, $w_m f(d) = 0$, where d is fibre density.

On the other hand, as described in section 3.3.1, matrix feedback is estimated to be $w_m = 0.27$ for an aligned matrix produced over seven days. This can be defined mathematically by saying that at the end of day 7, $w_m f(d) = 0.27$, where d is fibre density. Using these two way-points at day 4 and day 7, simulations were run computing average fibre density across the entire matrix at different times. Fibre density d was taken to be the number of fibres in the bin selected for matrix guidance in each matrix grid point. Simulations showed that the average fibre densities could be defined as

$$\bar{d} = \begin{cases} 5, \text{ at day } 4, \\ 10, \text{ at day } 7. \end{cases} \quad \textbf{(34)}$$

These two way-points were then used to define a linear function with conditions $f(5) = 0$ and $f(10) = 1$, where w_m would then be a user-defined maximal level of matrix feedback. This function can therefore be written as:

$$f(d) = \begin{cases} 0, & \text{if } d \leq 5, \\ \dfrac{d-5}{5}, & \text{if } 5 \leq d \leq 10, \\ 1, & \text{if } d \geq 10. \end{cases} \quad \textbf{(35)}$$

3.5.2 Motivation for modelling fibre organisation

In the previous simulations in section 3.4, cells had simply deposited a single fibre to the underlying matrix fibres at each time step, oriented in the direction of cell migration. Fibres were not reoriented by fibroblasts and deposition parameter (o_{dep}) had been fixed whilst the weighting of matrix feedback had been varied in order explore the effect of matrix feedback. Instead now in simulations matrix fibres could

be organised in three different ways: cells could increase the number of fibre elements, cells could realign existing fibre elements, and cells could degrade existing filament elements. This manipulation of the second layer of the model was employed to see how dynamic remodelling of the ECM might affect matrix patterns by mapping new points in metric space onto the original PCA given in Figure 51. The original points from Figure 51 are shown faintly on Figure 54 - Figure 57.

3.5.3 Altering the number of fibres deposited by fibroblasts

First, the effect of altering the number of fibres deposited by a fibroblast in a single time step (o_{dep}) was explored by fixing the rearrangement parameter (o_{re}) to zero, degradation parameter $o_{dep} = 1$, and matrix feedback to $w_m = 0.04$. Four different conditions were considered: Cells with low migratory noise and low fibre deposition ($o_{dep} = 2$), (Figure 54A, light orange circle); low migratory noise and high fibre deposition ($o_{dep} = 10$) (Figure 54A, dark orange circle); high migratory noise and low fibre deposition ($o_{dep} = 2$) (Figure 54A, light blue circle); high migratory noise and high fibre deposition ($o_{dep} = 10$) (Figure 54A, dark blue circle). Example matrices produced under these four conditions is shown in Figure 54B with starplots characterising the different matrix patterns. These four conditions represent extremes in parameter space. An increase in deposition results in higher HDM and higher curvature. Further, higher values of o_{dep} causes lower LRA as the matrix becomes so dense that there is a loss of order. The same analysis done with high matrix feedback ($w_m = 0.2$) showing very similar results (Figure 55).

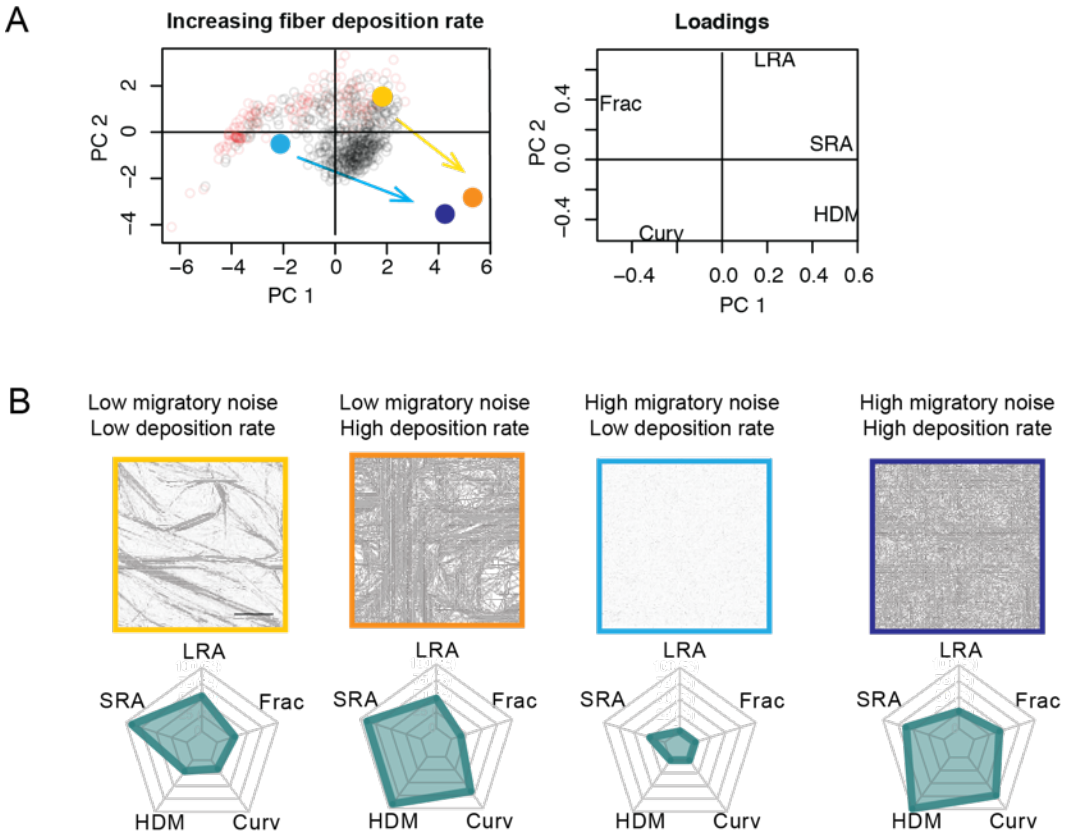


Figure 54: Changing deposition parameter (o_{dep}) with low matrix feedback

Analysing the effects of varying deposition parameter ($\boldsymbol{o_{dep}}$) and noise when matrix feedback is low ($w_m = 0.04, w_c = 0.03$). (A) PCA for aligning cells with low deposition (light orange circle, $\eta = 0, o_{dep} = 2, o_{deg} = 1, o_{re} = 0$), aligning cells with high deposition (dark orange circle, $\eta = 0$, $o_{dep} = 10, o_{deg} = 1, o_{re} = 0$), non-aligning cells with low deposition (light blue circle, $\eta = 0.14, o_{dep} = 2, o_{deg} = 1, o_{re} = 0$) and non-aligning cells with high deposition (dark blue circle, $\eta = 0.14, o_{dep} = 10, o_{deg} = 1, o_{re} = 0$). Blue arrow indicates change in deposition parameter ($\boldsymbol{o_{dep}}$) for non-aligning cells, yellow indicates change in deposition parameter ($\boldsymbol{o_{dep}}$) for aligning cells. Background points and loadings are from Figure 51. (B) Corresponding example stills of the matrix produced by different conditions and their starplots. N=10 simulations per point in parameter space.

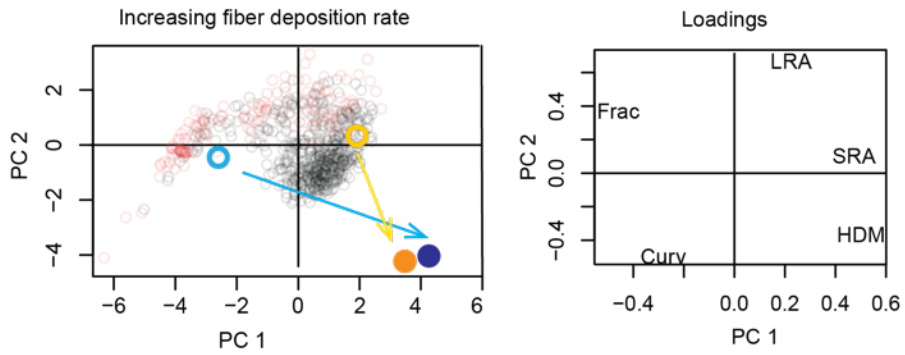


Figure 55: Changing deposition parameter (o_{dep}) with high matrix feedback

Analysing the effects of varying fibre deposition parameter ($\boldsymbol{o_{dep}}$) when matrix feedback is high ($w_m = 0.2, w_c = 0.03$). (A) PCA for aligning cells with low deposition (light orange circle, $\eta = 0, o_{dep} = 2, o_{deg} = 1, o_{re} = 0$), aligning cells with high deposition (dark orange circle, $\eta = 0$, $o_{dep} = 10, o_{deg} = 1, o_{re} = 0$), non-aligning cells with low deposition (light blue circle, $\eta = 0.14, o_{dep} = 2, o_{deg} = 1, o_{re} = 0$) and non-aligning cells with high deposition (dark blue circle, $\eta = 0.14, o_{dep} = 10, o_{deg} = 1, o_{re} = 0$). Blue arrow indicates change in ($\boldsymbol{o_{dep}}$) for non-aligning cells, yellow indicates change in ($\boldsymbol{o_{dep}}$) for aligning cells. Background points and loadings are from Figure 51. (B) Corresponding example stills of the matrix produced by different conditions and their starplots. N=10 simulations per point in parameter space.

3.5.4 Altering the number of fibres rearranged by fibroblasts

Next, the number of fibres which fibroblasts could reorganise in a single time step (o_{re}) was varied. The deposition and degradation parameters were fixed to $o_{dep} = 1, o_{deg} = 0$, and matrix feedback at $w_m = 0.04$. Again four different conditions were considered, each representing extremes in parameter space: Cells with low migratory noise and low fibre rearrangement ($o_{re} = 0$) (Figure 56A, light orange circle); low migratory noise and high fibre rearrangement ($o_{re} = 10$) (Figure 56A, dark orange circle); high migratory noise and low fibre rearrangement ($o_{re} = 0$) (Figure 56, light blue circle); high migratory noise and high fibre rearrangement ($o_{re} = 10$) (Figure 56, dark blue circle). Example matrices produced under these four conditions is shown in Figure 56B with starplots characterising the different matrix patterns.

Interestingly, a higher rearrangement value for o_{re} leads to fibres being organised into thicker bundles, leading to higher HDM especially for cells with high migratory noise. The matrix images produced *in silico* show the density of the most recently chosen bin at each grid point. The effects of matrix reorganisation remain largely constant regardless of the level of matrix feedback. This suggests that the effect of matrix homeostasis on cell organisation is subtler than the absolute level of matrix feedback. These results highlight the capacity for the model to evaluate the effects of matrix organisation on pattern generation and highlight how the introduction of dynamic remodelling increases matrix diversity by increasing fibre bundling. The

same analysis run with a high matrix feedback level of $w_m = 0.2$ shows very similar results (Figure 57).

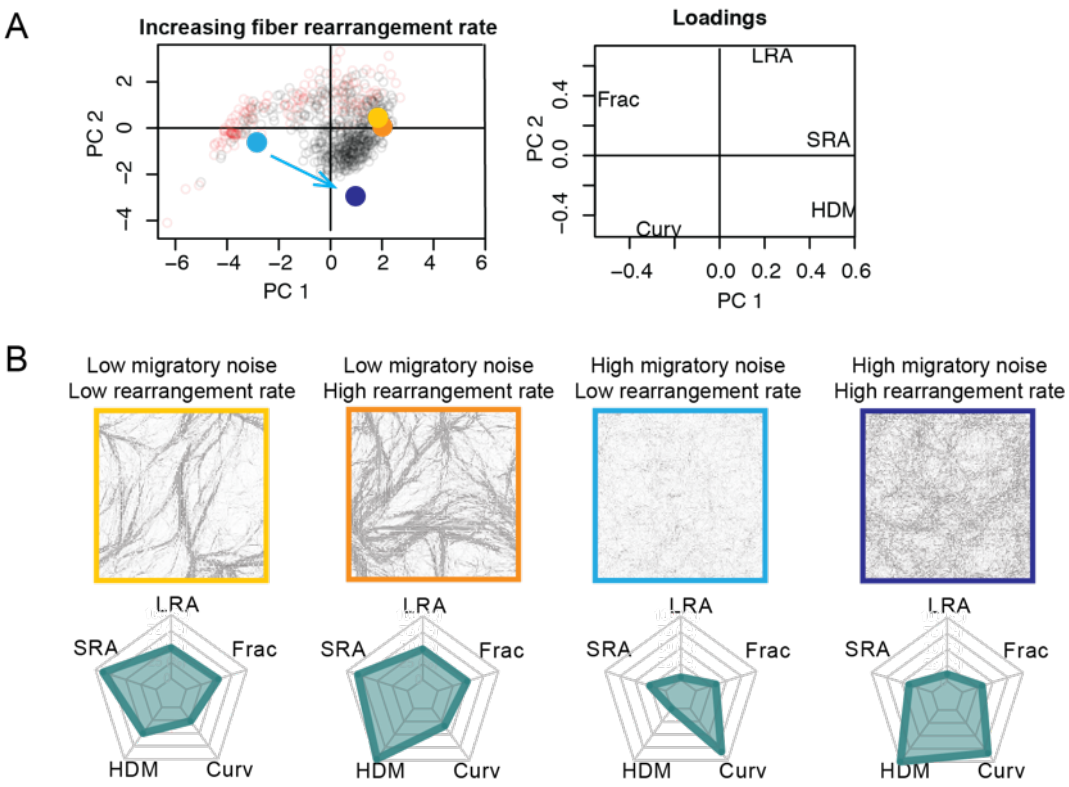


Figure 56: Changing rearrangement parameter with low matrix feedback

Analysing the effects of varying fibre rearrangement when matrix feedback is low ($w_m = 0.04$). (A) PCA for aligning cells with low rearrangement (light orange circle, $\eta = 0$, $o_{dep} = 1, o_{deg} = 0, o_{re} = 0$), aligning cells with high rearrangement (dark orange circle, $\eta = 0$, $o_{dep} = 1, o_{deg} = 0, o_{re} = 10$), non-aligning cells with low rearrangement (light blue circle, $\eta = 0.14$, $o_{dep} = 1, o_{deg} = 0, o_{re} = 0$) and non-aligning cells with high rearrangement (dark blue circle, $\eta = 0.14$, $o_{dep} = 1, o_{deg} = 0, o_{re} = 10$). Blue arrow indicates change in rearrangement (o_{re}) for non-aligning cells, yellow indicates change in rearrangement (o_{re}) for aligning cells. Background points and loadings are from Figure 51. (B) Corresponding example stills of the matrix produced by different conditions and their starplots. N=10 simulations per point in parameter space.

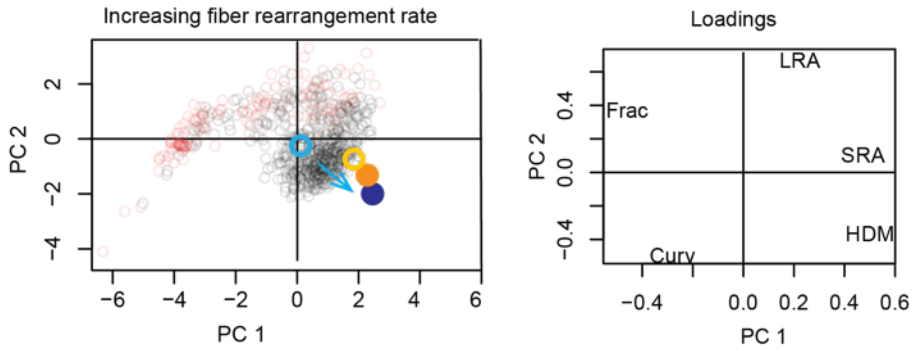


Figure 57: Changing rearrangement parameter with high matrix feedback

Analysing the effects of varying fibre rearrangement when matrix feedback is low ($w_m = 0.2$). (A) PCA for aligning cells with low rearrangement (light orange circle, $\eta = 0$, $o_{dep} = 1, o_{deg} = 0, o_{re} = 0$), aligning cells with high rearrangement (dark orange circle, $\eta = 0$, $o_{dep} = 1, o_{deg} = 0, o_{re} = 10$), non-aligning cells with low rearrangement (light blue circle, $\eta = 0.14$, $o_{dep} = 1, o_{deg} = 0, o_{re} = 0$) and non-aligning cells with high rearrangement (dark blue circle, $\eta = 0.14$, $o_{dep} = 1, o_{deg} = 0, o_{re} = 10$). Blue arrow indicates change in rearrangement (o_{re}) for non-aligning cells, yellow indicates change in rearrangement (o_{re}) for aligning cells. Background points and loadings are from Figure 51. (B) Corresponding example stills of the matrix produced by different conditions and their starplots. N=10 simulations per point in parameter space.

A limitation of the model rests in the output depiction of matrix fibres. In Figure 56, cells with low migratory noise produce thicker bundles of fibres deposited in the same bin, despite having the same overall fibre value for o_{dep} as cells with high migratory noise (Figure 56, yellow and light blue boxes). This is because the output visual matrix shows the density of the matrix in the dominant bin. For example, consider a matrix grid point which has ten fibres deposited across its eight bins. If all fibres have been deposited in bin 1, then the visual output matrix at this point has the orientation of bin 1 and intensity 10. If on the other hand the fibres are deposited with more dispersion by fibroblasts with very high migratory noise then bin 1 could have three fibres and all the other bins could have a single fibre in them. In this case the visual output matrix at this point has the orientation of bin 1 but with intensity 3. Clearly this is not optimal and needs further development, but is still sufficient to see overall patterning.

Table 7: Metric values for simulations with differing fibre organisation

	LRA	SRA	HDM	Curv	Frac

Figure 54 (yellow)	0.55	0.9	0.21	13	1.165
Figure 54 (orange)	0.46	0.89	0.91	30	1.111
Figure 54 (light blue)	0	0.24	0	NA	NA
Figure 54 (dark blue)	0.27	0.62	0.99	32	1.216
Figure 55 (yellow)	0.32	0.87	0.27	14	1.203
Figure 55 (orange)	0.13	0.72	0.82	33	1.132
Figure 55 (light blue)	0.02	0.24	0	NA	NA
Figure 55 (dark blue)	0.17	0.73	0.95	30	1.066
Figure 56 (yellow)	0.48	0.87	0.2	18	1.259
Figure 56 (orange)	0.48	0.82	0.39	20	1.242
Figure 56 (light blue)	0.04	0.26	0	35	1.181
Figure 56 (dark blue)	0.11	0.38	0.4	35	1.179
Figure 57 (yellow)	0.26	0.82	0.24	19	1.212
Figure 57 (orange)	0.24	0.8	0.41	23	1.183
Figure 57 (light blue)	0.25	0.63	0.01	25	1.28
Figure 57 (dark blue)	0.27	0.79	0.5	30	1.196

3.5.5 Exploring the effects of long-range contractility

Whilst the model does not explicitly incorporate contractile forces, it is well documented that fibroblasts are indeed able to pull on and hence rearrange fibres over a distance greater than the immediate area containing the cell (Kaur *et al.*, 2019). It is therefore important to understand the impact of long-range contractility in matrix patterning in particular, to establish whether there exists an optimal level of contractility for ECM alignment. The model was extended to include a parameter for rearrangement length (o_{length}), indicating the distance over which fibres could be reoriented. Simulations were run with fixed noise, cell-cell guidance and zero matrix feedback ($\eta = 0, w_c = 0.03, w_m = 0$) varying rearrangement length and

rearrangement rate. Cells were able to reorient fibres within the region that coincided with the cell's long axis (Figure 58). These two terms were then varied in simulations with cells with either low individual migratory noise (favouring alignment) or high migratory noise (favouring no alignment). The low migratory noise simulations are shown in the upper panels and the high migratory noise in the lower panels. LRA over a seven-day time-course is shown to give an impression of the rate of emergent alignment. Very high values of o_{re} and intermediate values of o_{length} were able to boost the rate of alignment for low noise fibroblasts (highlighted with green box), and led to a small gain in alignment for the higher noise fibroblasts (highlighted with grey box). These data indicate that remodelling of ECM at a distance can influence alignment, but the magnitude of the effect is small. (Figure 59).

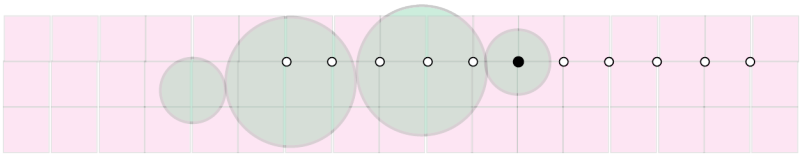


Figure 58: Schematic of long-range fibre organisation

The cell is associated with a grid point (shown with a black dot), which is the grid point closest to the centre of the cell's head. In the event that $w_m \neq 0$, the cell will be guided by the fibres in that grid point. The cell will also reorganise fibres in that grid point. When long-range rearrangement of fibres is implemented ($o_{length} \neq 0$), the cell can also rearrange fibres in adjacent boxes along the direction of the cell's long axis (white dots). In this schematic $o_{length} = 5$.

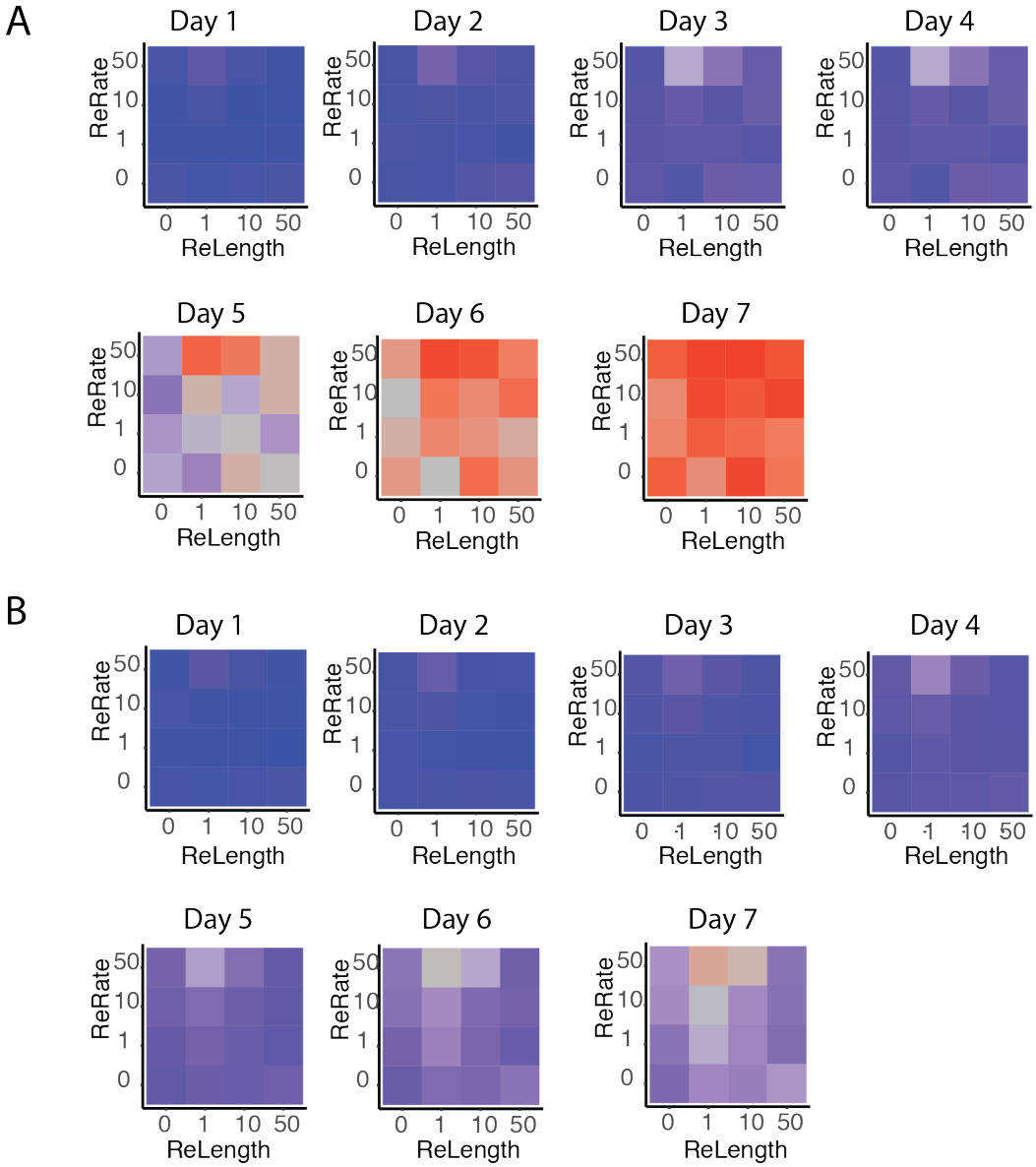


Figure 59: Long-range contractility

Simulations varying length over which matrix fibres can be rearranged by fibroblasts (o_{length}) and the number of fibres rearranged by fibroblasts in a single time step (o_{re}). Red indicates a high alignment score and blue a low alignment score. (A) Heatmaps showing how long-range alignment (LRA, $0 - 400\mu m$) changes over time for simulations of cells with low migratory noise. (B) Heatmaps showing how long-range alignment (LRA) changes over time for simulations of cells with high migratory noise

3.5.6 Model construction as a tool for challenging biological assumptions

In the construction of the model, it is necessary to make certain assumptions, which are made based on a compromise between mathematical tractability and

biological justification. The topic of fibre rearrangement is a prime example of how model development can answer existing questions as well as generating new questions.

Constructing rules for long-range contractility has led us to ask the following questions:

1. Would fibres perpendicular to the cell be rearranged? Or would only fibres in neighbouring bins be rearranged?
2. What is the timescale for fibre rearrangement to happen? Would fibres more closely aligned be rearranged more quickly?
3. If two different fibroblasts pull on the same fibre, what is the effect? Does the effect of the nearest fibroblast dominate? What is the net average effect?
4. What is the effect of altering the neighbourhood of long-range contractility? Is overall patterning affected if cells can alter fibres not only coinciding with their long axis of orientation?

These are exciting areas of future research, and lend themselves to being understood by switching iteratively between experimental and computational work. These challenging questions certainly cannot be answered without mathematical modelling. It is clear that these questions lead to a more force-based paradigm than the rule-based system of the current model and other models more equipped to dealing with forces may be more appropriate to answer these questions of how exactly fibroblasts pull on fibres, which recently begun to be elucidated experimentally (Steinwachs *et al.*, 2015). Incorporated data on cell-matrix forces such as traction force microscopy will provide more thorough understanding on the effects of fibre rearrangement on emergent matrix patterning (3.5.4), but may not easily translate to a rule-based system such as the one presented here. However, with the current lack of information on how these forces work on the mesoscale, agent-based modelling provides a good theoretical framework for exploring the interactions between cells and fibres. In the building of such models, one is forced to consider what biological assumptions are appropriate, and in doing so, is led to propose new questions from a unique perspective.

3.6 Mimicking *in vivo* tissues

The model has so far demonstrated that a wide range of matrix patterns can be generated by varying a small number of parameters. As shown in Figure 2, diverse patterning is observed in both normal and pathological tissues *in vivo*. It is currently unknown how patterns such as these arise *in vivo* across a mesoscale of up to several millimetres however, the fibroblasts producing such patterns will most likely have different functional properties and lineages (Driskell *et al.*, 2013). It is of interest therefore to establish if the model can mimic such *in vivo* patterns and with what parameters this might be possible ie what fibroblast/matrix properties can produce such patterns. Understanding which parameters could produce these *in vivo* patterns would shed insight into how these patterns arise through cell and matrix interactions across many orders of magnitude.

The model was challenged to mimic the *in vivo* tissues of the dermis, liver, spleen and stomach. To do this, several alterations were made to the simulations. Firstly, the matrix grid was made more fine-grained, consisting now of 256x256 grid points as opposed to 128x128, which was used to previous simulations. This allowed for higher precision of fibre deposition, at the cost of computational speed. Secondly, by counting nuclei in the *in vivo* images from Figure 32, it was deduced that N=200 cells was an appropriate level of confluence. Thirdly, simulations were run for three times longer than normal simulations, reflecting the gradual development of these tissues over long time periods.

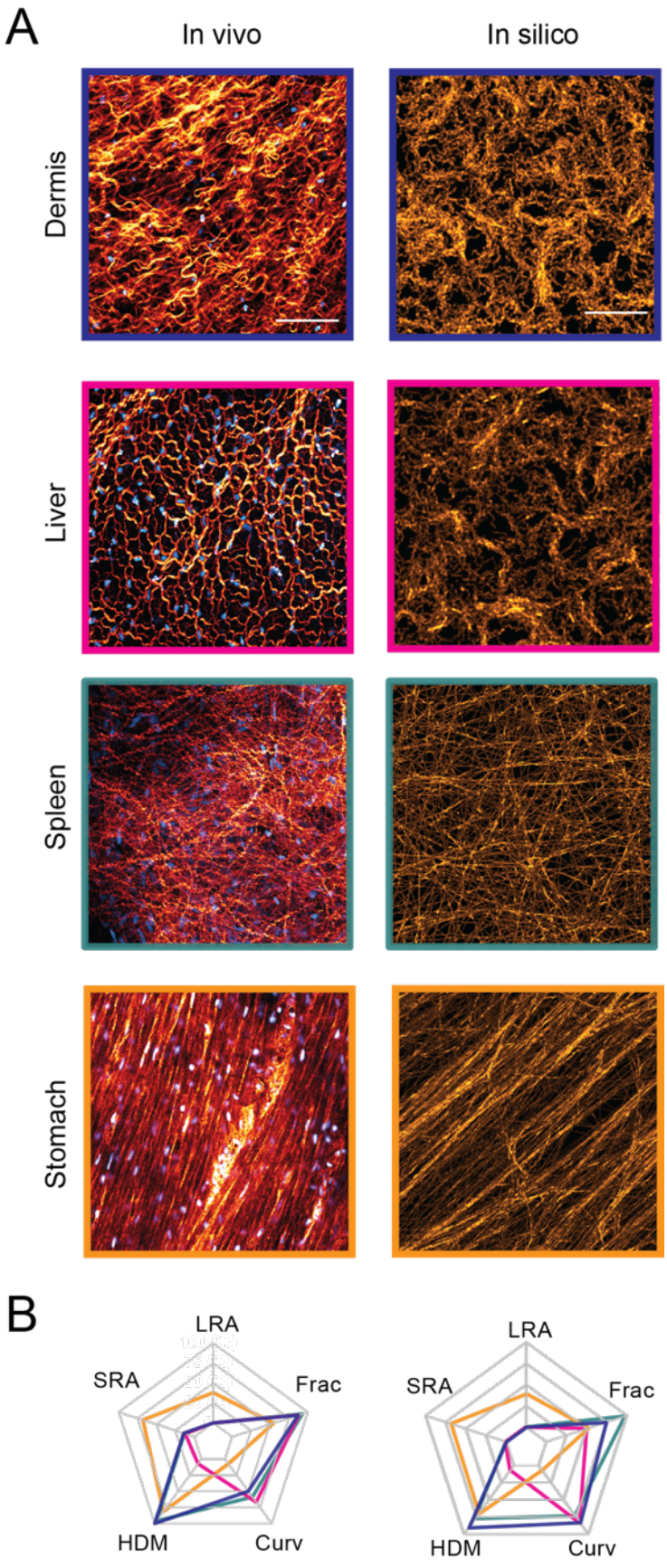


Figure 60: Mimicking *in vivo* tissues *in silico*

(A) Second harmonic imaging of collagen (orange) in various tissues from a PDGFR nuclear labelled mouse (left column). PDGFR positive nuclei (cyan) indicates fibroblast density. *In silico* matrix that mimics the *in vivo* patterns are shown in the right column. (B) Each matrix from part (A) is characterised on a starplot in a corresponding colour, to long-range alignment (LRA), short-range alignment (SRA), percentage of high-density matrix (HDM), curvature (Curv) and fractal dimension (Frac). Differences in parameters for generating such matrix patterns are given in Table 8. Scale bars represent 100μm.

With these modifications, simulations were run stepping incrementally through parameter space. In section 2.7.1, it was established that *in vitro*, non-aligning fibroblasts moving on glass in sub-confluence have noise attuned to $\eta = 0.18$, corresponding to very high individual migratory noise. From this it was deduced that searching in the range $\eta \in [0, 0.2]$ would be an appropriately wide range of values. In section 2.6.2, the fitted value of cell-cell collision guidance for the aligning fibroblasts was ascertained experimentally with a given value of $w_c \approx 0.05$. Therefore, the range $w_c \in [0,0.2]$ was also chosen as a reasonably wide range of values to search. In section 3.3.1 it was established that matrix feedback *in vitro* took an approximate value of $w_m \approx 0.27$. This level of matrix feedback is in two dimensions, whereas in three dimensions *in vivo* matrix feedback is likely to be higher given the apical and basal constraints that matrix fibres can provide to cells. Therefore, the range for w_m and kept to a range of $w_m \in [0,1]$. Noise and cell-cell guidance were varied by increments of 0.01 for their specified ranges. Matrix feedback was varied by increments of 0.01 in the range [0,0.1] and by increments of 0.1 in the range [0.1,1]. Simulations were run for all combinations of values within these ranges. In this way, simulations were run through all of parameter space.

The three principal parameters of individual migratory noise, cell-cell collision guidance and matrix feedback, together with cell speed and fibre deposition per time step (o_{dep}) were varied to try to recapitulate the *in vivo* tissues. In order to choose the most suitable areas in parameter space to generate patterns similar to the *in vivo* tissues, a number of heuristic arguments were made:

The stomach is highly aligned. From stepping through parameter space, the model shows that the area of parameter space corresponding to high LRA and SRA is where noise and matrix feedback are low, and cell-cell guidance is non-zero.

The *in vivo* liver is highly corralled which the model shows to occur in the region of parameter space where matrix feedback is high.

The spleen matrix appears to be made up of long straight fibres, suggesting low individual migratory noise, which are uncoordinated, suggesting low cell-cell guidance and independent, suggesting low matrix feedback.

The dermis matrix is swirl-like, diffuse and reinforced, which as established in 3.4.2 corresponds to the region of parameter space where matrix feedback and noise are high.

Through this line of argument, parameters were chosen to generate the *in silico* patterns shown in Figure 60 and Table 8. Qualitatively they agree well with the *in vivo* images.

Table 8: Parameter values corresponding with Figure 60

	Dermis	Liver	Spleen	Stomach
Noise (η)	0.1	0.1	0.01	0.0
Cell-cell guidance (w_c)	0.03	0.03	0	0.03
Matrix feedback (w_m)	0.4	0.8	0.01	0
Deposition rate	2	1	2	2
Mean speed (μ/h)	0.7	0.7	1.8	3.5
Speed standard deviation (μ/h)	0.1	0.1	0.3	0.5
Number of matrix grid points	256^2	256^2	256^2	256^2

3.6.1 Comparing *in vivo* and *in silico* metrics

To compare how well the *in silico* matrix patterns agreed with the *in vivo* patterns, the five metrics of LRA, SRA, HDM, Frac and Curv were derived. In both the *in vivo* and *in silico* patterns, the long-range alignment was now computed at a distance of

$200\mu m - 800\mu m$ and short-range alignment was computed at $0\mu m - 200\mu m$ to account for different size and resolution of the matrix images. The raw values for the metrics are given below in Table 9:

Table 9: Metric values corresponding to Figure 60

	LRA	SRA	HDM	Curv	Frac
IN VIVO					
Dermis	0	0.14	1	40	1.549
Liver	0	0.12	0.54	47	1.554
Spleen	0	0.13	0.98	44	1.571
Stomach	0.38	0.68	0.91	23	1.42
IN SILICO					
Dermis	0	0.11	0.32	37	1.434
Liver	0	0.16	0.02	37	1.411
Spleen	0.15	0.19	0.27	34	1.458
Stomach	0.99	0.99	0.25	14	1.412

An important challenge arises from these values: there are discrepancies between the *in vivo* and *in silico* metrics. The stomach matrix *in vivo* looks highly aligned, and yet has LRA=0.38 and SRA=0.68. An *in silico* matrix that looks similarly aligned would have much higher LRA and SRA, closer to 1. This is because *in silico*, the output fibre orientations are recorded exactly, whereas *in vivo* the images are noisier with variation in fibre intensity. Whilst the theoretical upper limit for both LRA and SRA is 1, the upper limit *in vivo* is probably closer to 0.38 and 0.68 respectively. Similarly, the mean value of HDM *in silico* through all parameter space is 0.3, and very few simulations have HDM higher than 0.5 whilst Frac *in silico* is also consistently lower than the four *in vivo* images. This is most likely due to differences in fibre intensity as picked up by the microscope, fibre thickness and duration over which the matrix patterns develop *in vivo*. Whilst the matrix patterns *in silico* qualitatively look very similar to *in vivo* patterns, there is a gap between the *in vivo* and *in silico* values due to the nature of how the matrix patterns are captured. Therefore, it is difficult to ascertain an appropriate function for “inter”-translation from *in silico* to *in vivo* values.

Encouragingly, whilst the raw measurements for the five metrics are different from the *in vivo* values for the reasons described above, the “intra”-comparison between the four patterns is consistent with the *in vivo* “intra”-comparisons. For example, the HDM of the liver both *in vivo* and *in silico* is the lowest of the four, the Frac and Curv of the stomach is the lowest of the four and the Frac of the spleen is the highest. This suggests that the areas in parameter space which produce the similar patterns relative to the system (*in silico* vs *in vivo*) have been correctly identified. Further evidence of this is given by normalising the *in vivo* and *in silico* values with the following functions shown in Table 10:

Table 10: Normalisation functions corresponding to Figure 60

Metric	Normalisation function *in vivo*	Normalisation function *in silico*
LRA	x	$\max(0, x - 0.6)$
SRA	x	$\max(0, x - 0.3)$
HDM	$\frac{x - 0.5}{0.5}$	$\frac{x}{0.35}$
Curv	$\frac{x - 20}{40}$	$\frac{x - 12}{30}$
Frac	$\frac{x - 1.2}{0.4}$	$\frac{x - 1.36}{0.1}$

These normalised values are then plotted onto starplot shown in Figure 60. Again, these show strong agreement between the *in vivo* patterns and the *in silico* patterns mimicking the *in vivo* case. It is exciting to note how it possible to recapitulate *in vivo* matrix patterns through a model with few parameters and simple rules. This offers a possible explanation for how such complex tissues arise in nature.

3.7 Interconversion between matrix types

So far, the model has been used to understand how patterns arise *de novo*. In other words, starting with nothing, how can patterns emerge? However, in many instances this is an oversimplified view and in fact new matrix patterns emerge

from existing matrix. The organisation of the ECM can change over time, either as the result of ageing, wound healing or more malign processes such as tumourigenesis.

3.7.1 Matrix transitions in cancer

As discussed in the Introduction, matrix organisation becomes dysregulated in cancer. In the tumour microenvironment, there are typically more fibroblasts than in normal tissue, many of which are CAFs, which display a highly activated phenotype. This leads to more matrix fibre proteolysis, degradation and rearrangement. Matrix interconversion in cancer was explored experimentally. Gomori trichome staining of collagen in a normal region and in a melanoma region of the dermis highlights the conversion process (Figure 61). Matrix in the normal region is characterised by curved matrix fibres. In contrast, the adjacent region of the dermis containing melanoma displays altered matrix organisation with more aligned fibres near to the melanoma. Additionally, changes in the matrix are observed in cancerous regions over time. Consistent with a previous report that targeted BRAF inhibition in melanoma activates stromal fibroblasts (Hirata *et al.*, 2015), a comparison of ECM organisation in the same patient prior to therapy and post-therapy indicated that when the treatment is failing, matrix alignment increases and curvature and fractal dimension decrease (Figure 62).

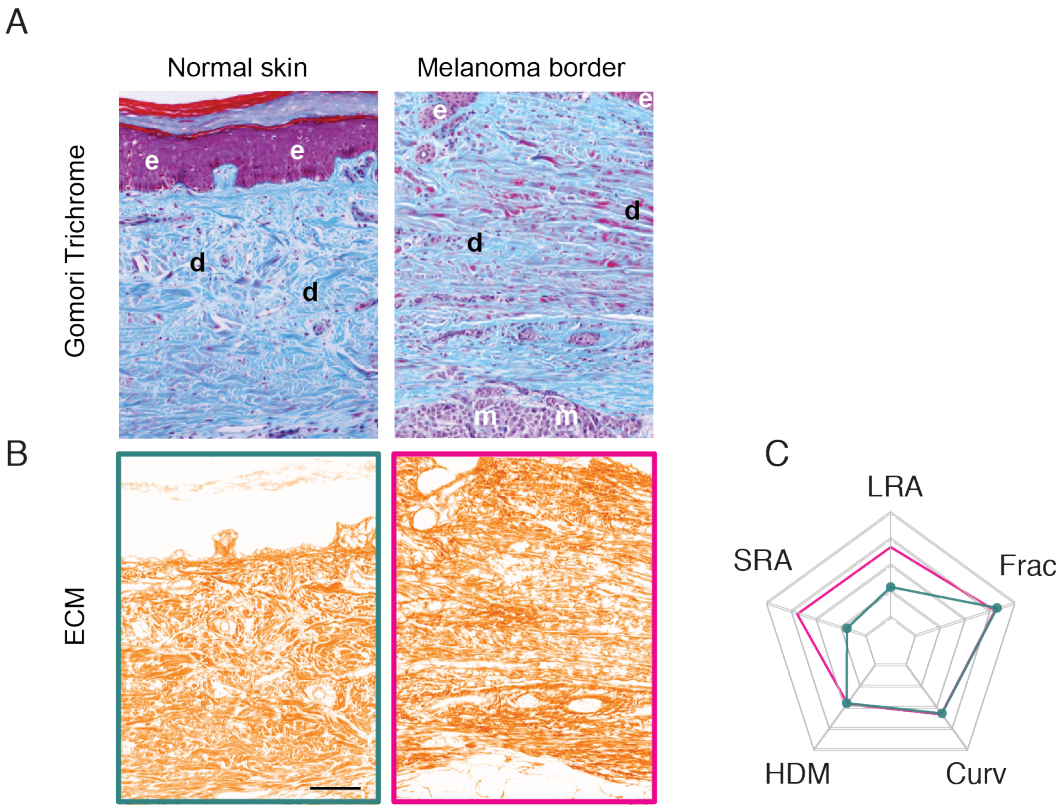


Figure 61: ECM organisation changes near the melanoma border

(A) Gomori Trichrome staining of normal skin (left) and the melanoma border (right). Epidermis is denoted with an “e”, dermis with a “d” and melanoma with an “m”. (B) Staining showing the collagen alone. (C) Corresponding starplot characterising normal dermis (green line) and melanoma (pink line). Scale bars represent $100\mu m$.

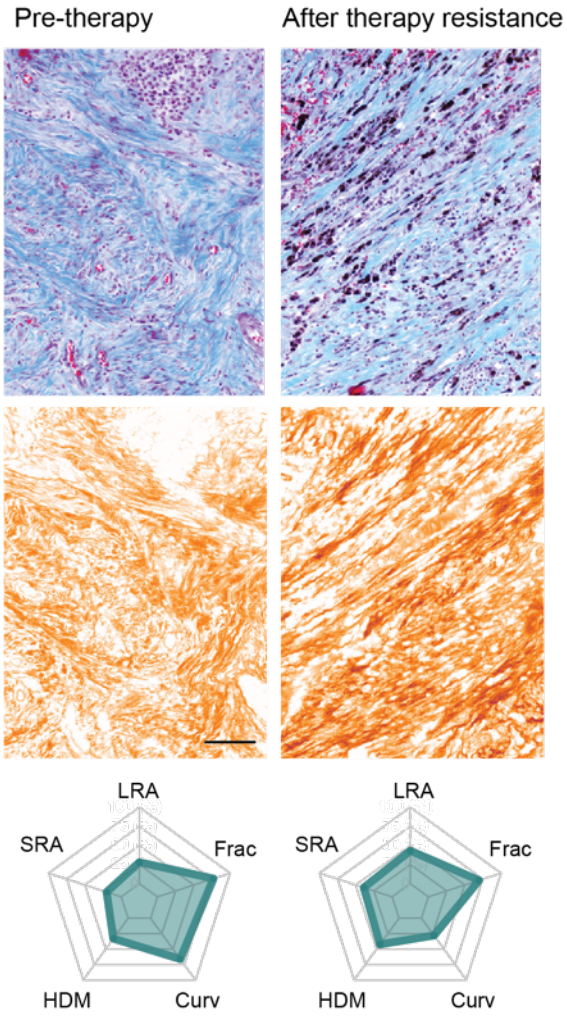


Figure 62: ECM organisation changes in therapy resistance

Gomori Trichrome staining showing from a sample from a melanoma patient pre-therapy and post-therapy when resistance has occurred. Matrix is shown in blue. Matrix alone is shown in orange below with corresponding starplots on the bottom row. Scale bars represent $100\mu m$

Analysis from Figure 50 suggests that high matrix feedback may preclude the ability for fibroblasts to remodel the matrix, as they would be forced to follow existing matrix. The model was employed to understand under what conditions such matrix interconversion might be observed, in particular what levels of matrix feedback would enable matrix remodelling. First preliminary simulations were run to generate “dermal”-like matrices, produced with high individual migratory noise and high matrix feedback to generate the normal pre-cancerous tissue. Fibroblasts with individual migratory noise reduced to zero were then reseeded at random on this dermal matrix. As expected, the model predicts that high levels of matrix feedback will lead to fibroblast migration being channelled into following existing matrix,

thereby preventing matrix remodelling (Figure 63B, blue line). However, lowering matrix feedback to 0 permitted a gradual transition from `dermal' matrix to a more aligned matrix (grey line). The emergent matrix showed hybrid features with retention of the original “dermal” matrix (Figure 63).

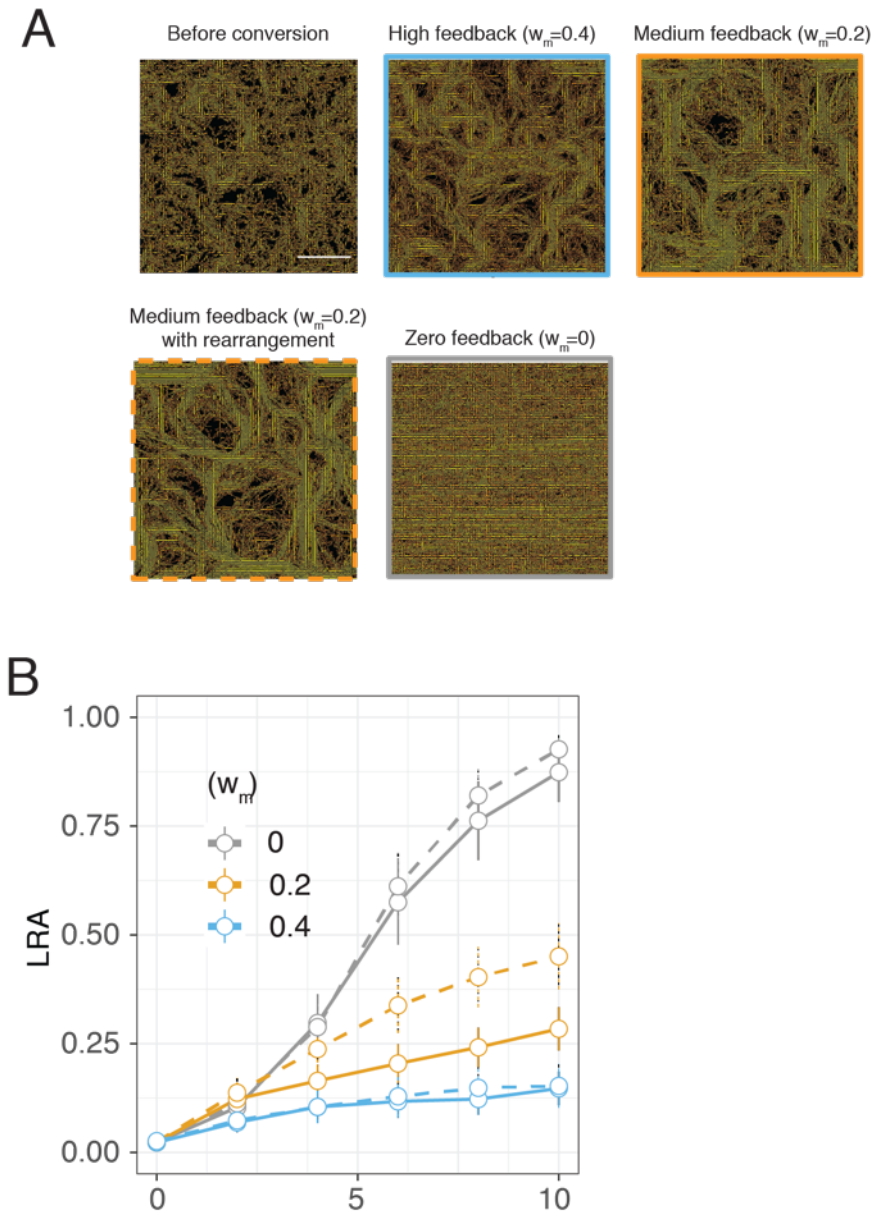


Figure 63: Mimicking the matrix transition in melanoma

Long-range alignment (LRA) over time beginning with *in silico* dermis for varying levels of matrix feedback: $w_m = 0$ (grey lines), $w_m = 0.2$ (orange lines), $w_m = 0.4$ (blue lines). $o_{dep} = 10$. Solid lines show simulations with degradation and rearrangement rates set to zero, whilst dotted lines show simulations with$o_{re} = 5, o_{deg} = 5$. At day 10, when matrix feedback $w_m = 0.2$, simulations with fibre degradation and rearrangement are more aligned (p=9e-04). N=25 simulations per

point in parameter space. (D) Corresponding *in silico* matrix showing before and after conversion. Colour brightness has been normalized across images. Scale bars correspond to $100\mu m$.

In order to reduce the artefact of the original matrix, similar simulations were next run with nonzero levels of matrix degradation and fibre rearrangement. The results indeed confirmed that at a middle level of matrix feedback ($w_m = 0.2$, Figure 63 orange line), enabling fibroblasts to rearrange and degrade fibres, produces a more aligned matrix (p=9e-04, two tailed t-test), suggesting that the ability of fibroblasts to degrade and rearrange fibres was key to the interconversion process.

To address the question of whether such a transition could occur experimentally, aligning fibroblasts were seeded on a thick non-aligned matrix and non-aligning fibroblasts on an aligned matrix, and observed over four days how the cells behaved on the matrix, and the patterning of the new matrix they produced (Figure 64). The original matrix is shown in yellow, with the emergent organisation of the fibroblasts at different times shown in phase-contrast below. The new matrix produced by these fibroblasts is shown in blue. In both cases the fibroblasts display an initial tendency to follow the original matrix (day 1), but then revert to their preferred phenotype (day 3-4), being able to ignore the original matrix and produce a new matrix on top. For completeness aligning fibroblasts were also seeded on aligned matrix and non-aligned fibroblasts were seeded on non-aligned matrix in separate experiments Figure 65. Together, the adapted two-layer nematic Vicsek model and supporting experimental analyses demonstrate that matrix feedback enables diverse emergent patterns of ECM, including curved matrix structure, but that the strength of matrix feedback, is not sufficient to lock the system indefinitely, therefore enabling transitions to occur over a timescale of days. Additionally, these results confirm how matrix feedback enables fibroblasts to take instruction from surrounding matrix (Dallon, Sherratt and Maini, 1999; McDougall *et al.*, 2006). Experimental methods are described in the Appendix 6.2.

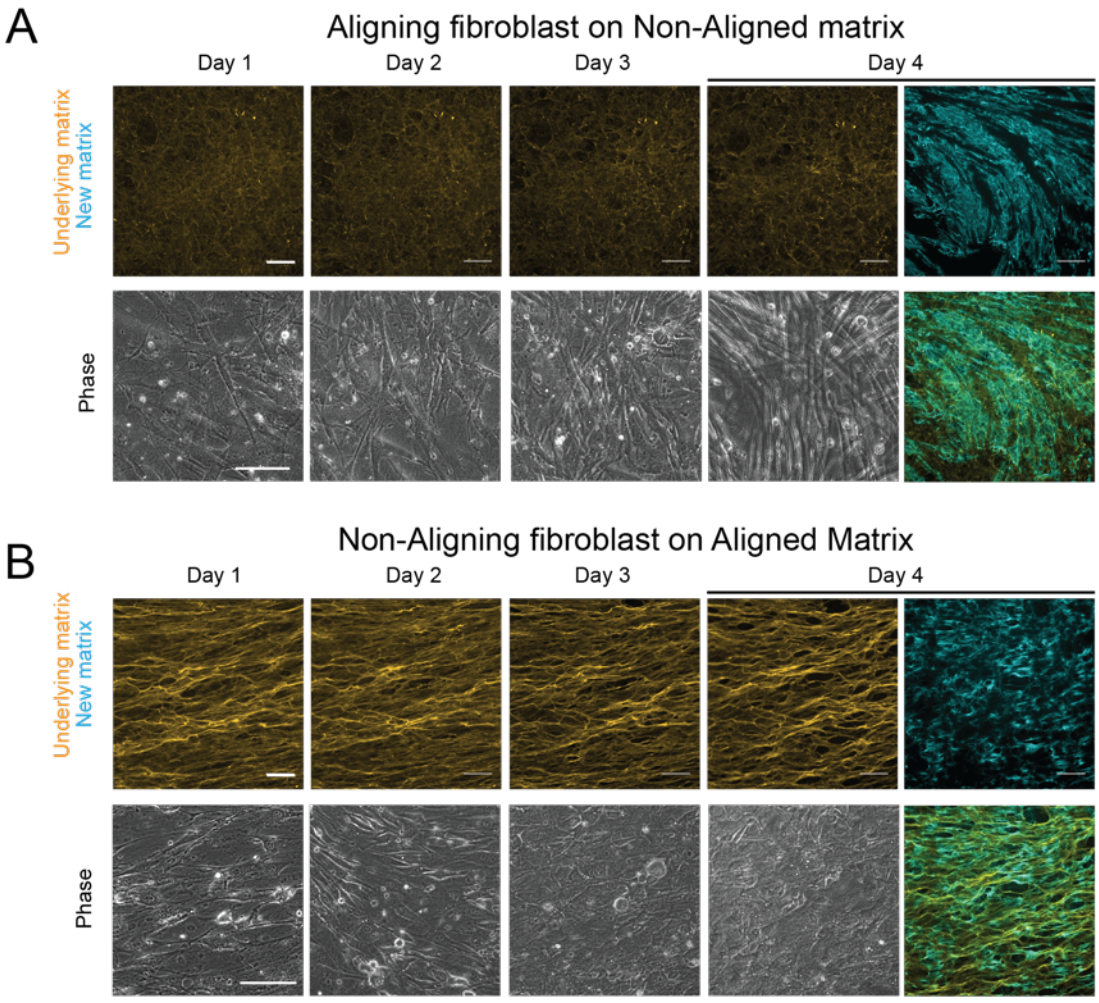


Figure 64: Seeding fibroblasts on opposite pre-existing matrix

Modelling conversion between matrix patterning *in vitro* (A) Non-aligning fibroblasts were seeded onto a pre-existing anisotropic matrix. The pre-existing matrix is shown in yellow, new matrix produced over the duration of the assay is shown in blue and the composite image after four days is shown on the right. Corresponding zoomed-in phase imaging of cell-body organisation below. For the first two days cells follow the matrix and align, but this order breaks down by day 3. (C) Aligning fibroblasts were seeded onto a pre-existing isotropic matrix. Corresponding zoomed-in phase imagine of cell-body organisation below. On day 1 cell body orientation is disorganised and follows the isotropic matrix. By day 3 fibroblasts start to ignore the matrix and align. Scale bars represent $100\mu m$.

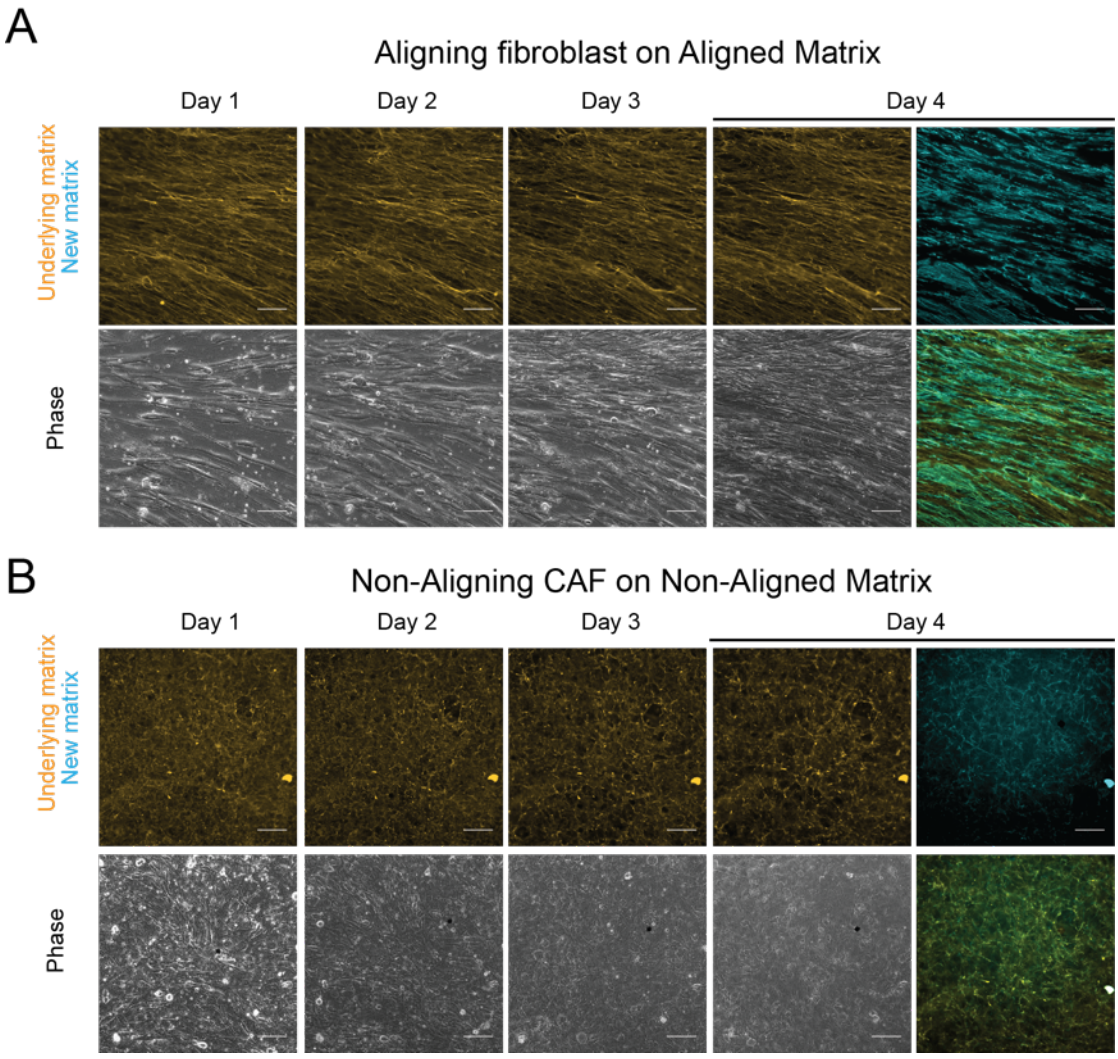


Figure 65: Seeding fibroblasts on matching pre-existing matrix

Complement experiments to those described in Figure 64. (A) Aligning fibroblasts were seeded onto a pre-existing anisotropic matrix. The pre-existing matrix is shown in yellow, new matrix produced over the duration of the assay is shown in blue and the composite image after four days is shown on the right. Corresponding zoomed-in phase imaging of cell-body organisation below. Fibroblasts quickly organise themselves to follow the existing ECM. (B) Non-aligning fibroblasts were seeded onto a pre-existing isotropic matrix. Corresponding zoomed-in phase image of cell-body organisation below. Scale bars represent $100\mu m$.

3.7.2 Matrix transitions in wound healing

Similar principles were then applied to explore regeneration of matrix after wounding (Figure 66). To simulate a wound, the model first generated matrix, then all cells were removed from the entire simulation space and all matrix from the wound area. Then fibroblasts were re-added at random in the region of ECM

outside the wound which would deposit fibres only within the wound area ($o_{dep} = 1$). For cell-types that generated aligned or swirly matrix, adding the same type of fibroblasts that created the original matrix proved ineffective at repairing the wound, leaving gaps and with the new matrix often not matching the old matrix (Figure 66, green box). An aligning fibroblast was designated "Phenotype A" ($\eta = 0, w_c = 0.03, w_m = 0$). Based on analysis in Figure 48 revealing that cells with high individual migratory noise and zero matrix feedback generate ECM that homogeneously covers the simulation space and is not porous, the 'wound repair' performance of these cells was tested. Figure 66 shows that this phenotype B ($\eta = 0.14,\ w_c = 0.03, w_m = 0$) could not produce similar matrix to before the wound (yellow box). Analysis of how well cells with varying levels of individual migratory noise and matrix feedback re-filled the space revealed that a higher level of individual migratory noise was essential for efficient restoration of matrix (Figure 66B). At day 2, phenotype A has filled 82% of the wound compared with 94% by phenotype B (two-sided t-test, p-value<2e-16, N=20). There was also a trend towards low levels of matrix feedback being beneficial. Mechanistically, high levels of matrix feedback made the movement of cells from areas of high ECM into low matrix unfavourable as it resulted in cells being channelled into following existing matrix.

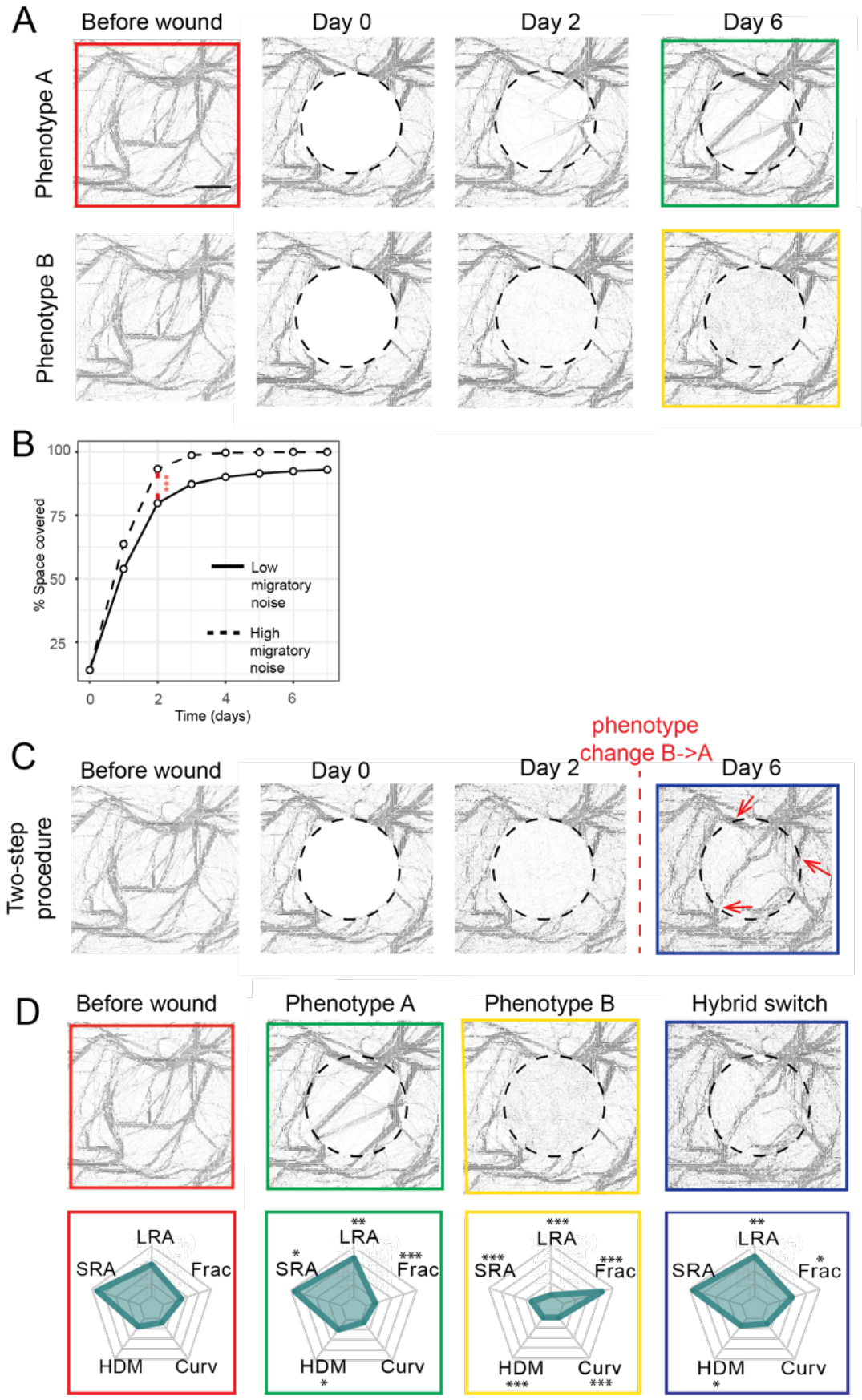


Figure 66: Simulating wound healing

(A) Wound healing simulation adding cells with high individual migratory noise and zero matrix feedback (top row Phenotype A, $\eta = 0, w_3 = 0.2$) and cells with low individual migratory noise and high matrix feedback (bottom row Phenotype B, $\eta = 0.14, w_3 = 0$). Outline of initial wound indicated by dotted line. (B) Graph indicating how wound is filled in by cells with low noise (solid line) and high noise (dotted line) when matrix feedback is set at $w_3 = 0.2$. For all simulations, number of cells = 800. N=20 simulations per point in parameter space. Red stars indicate p-value<0.001, two-sided t-test. (C) Wound healing simulation adding non-aligning cells with zero matrix feedback, which after two days revert to being like the original cell type, with low noise and high matrix feedback (days 0-2: Phenotype B, $\eta = 0.14, w_3 = 0$, days 3-7: Phenotype A, $\eta = 0.0, w_3 = 0.2$). (D) Starplots characterising each wound-repair option. Two-step procedure matches original matrix before the wound more closely (one-sided Mann-Whitney U test, two-step vs phenotype A, p=0.01, two-step vs phenotype B, p<2e-16. Wounding on five different matrices with N=20 simulations for each procedure type) Stars indicate p-values showing statistically clear relationships for each metric as compared to the pre-wound matrix. One star: p-value<0.05, two stars: p-value<0.01, three stars: p-value<0.001. Scale bars represent $100\mu m$.

To explore the implications of the above findings, it was next considered if wound repair to re-generate the same matrix that was initially present might require multiple stages of fibroblast phenotype activity. Specifically, if fibroblasts with one set of mechanistic properties was optimal for initially filling the space (Phenotype B) then a transition to a new set of mechanistic properties may enable remodelling of the matrix to re-generate the original pattern (Phenotype A). Figure 66C shows an example wound-healing scenario where fibroblasts display high individual migratory noise and zero matrix feedback for the first two days of healing (Phenotype A). After two days, they then revert back to having the same characteristics as the fibroblasts which produced the original matrix, with low individual migratory noise and high matrix feedback (Phenotype B). Comparing the example one-step wound healing process with the two-step process (blue box) shows that the two-step wound-healing process is more effective at filling in the wound and creating new matrix similar to before the wound. Interestingly, the model also shows a tendency for matrix feedback to induce the new matrix to be ‘connected’ to the existing matrix, resulting in the old matrix at the periphery of the wound being ‘stitched together’ (Figure 66C, zoomed-in blue box, example stitching indicated with red arrows. This suggests that it may be beneficial to induce matrix feedback to be stronger later on in wound repair, which could then facilitate more effective matrix repair by joining fibres together. The two-step procedure was ranked against phenotype A and phenotype B alone for each metric and found that it improved the

match between the new matrix and the original matrix before the wound (Figure Figure 66D, one-sided Mann-Whitney U test, phenotype A (green box) vs two-step (blue box) p=0.01, phenotype B (yellow box) vs two-step (blue box) p<2e-16).

3.7.3 Matrix transitions in ageing

Finally, the model was used to confirm how matrix of the dermis changes during the ageing process, in particular how the matrix is degraded with ageing Figure 67A. Matrix structure changes over time, comprising a decrease in fine diffuse fibres and an increase in more concentrated fibre bundles. This increase in discernible fibre bundles causes an increase in fractal dimension. The young dermis is characterised by a diffuse swirly matrix, with high HDM. The model predicts that a similar matrix can be produced *de novo* from cells with high noise and high matrix feedback. Such a matrix was generated with the model (Figure 67B, red box). The old dermis has lower HDM and has lost its diffusivity and the fibre bundles are more clearly defined (Figure 1c). The model predicts that such a matrix could be produced *de novo* from cells with low noise and high matrix feedback. In Figure 67, such swirly patterns typically have higher HDM than in the diffuse swirls, however the matrix of the old mouse has lower HDM than the young dermis. It was of interest to model the transition between the young and old dermis, specifically by reducing HDM, reinforcing thick fibre bundles, and degrading thinner fibres. It can be conjectured from the analysis in 3.4 that for the reduction in HDM, there would need to be matrix degradation. Additionally, to reinforce certain thick fibre bundles matrix feedback would need to be quite high and that deposition would need to be as high as degradation in these areas. Simulations were run varying matrix feedback and deposition parameter o_{dep} and fixing $o_{deg} = 1$. The best agreement of HDM between the old matrix *in vivo* and *in silico* occurred when matrix feedback is set to $w_m = 0.12$ and $o_{dep} = 1$. The final matrix *in silico* agrees qualitatively with that of the old mouse dermis (Figure 67, yellow box). Interestingly, matrix produced de novo with these parameters, but with the $o_{deg} = 0$ to allow matrix accumulation, did not generate the pattern observed in the aged dermis (shown in Figure 67, green box). These analyses demonstrate how when transitioning between two different matrix types, cells can take instruction from the

original matrix and produce a hybrid matrix that can have different patterning to de novo matrix generation and that this remodelling of matrix could be seen experimentally.

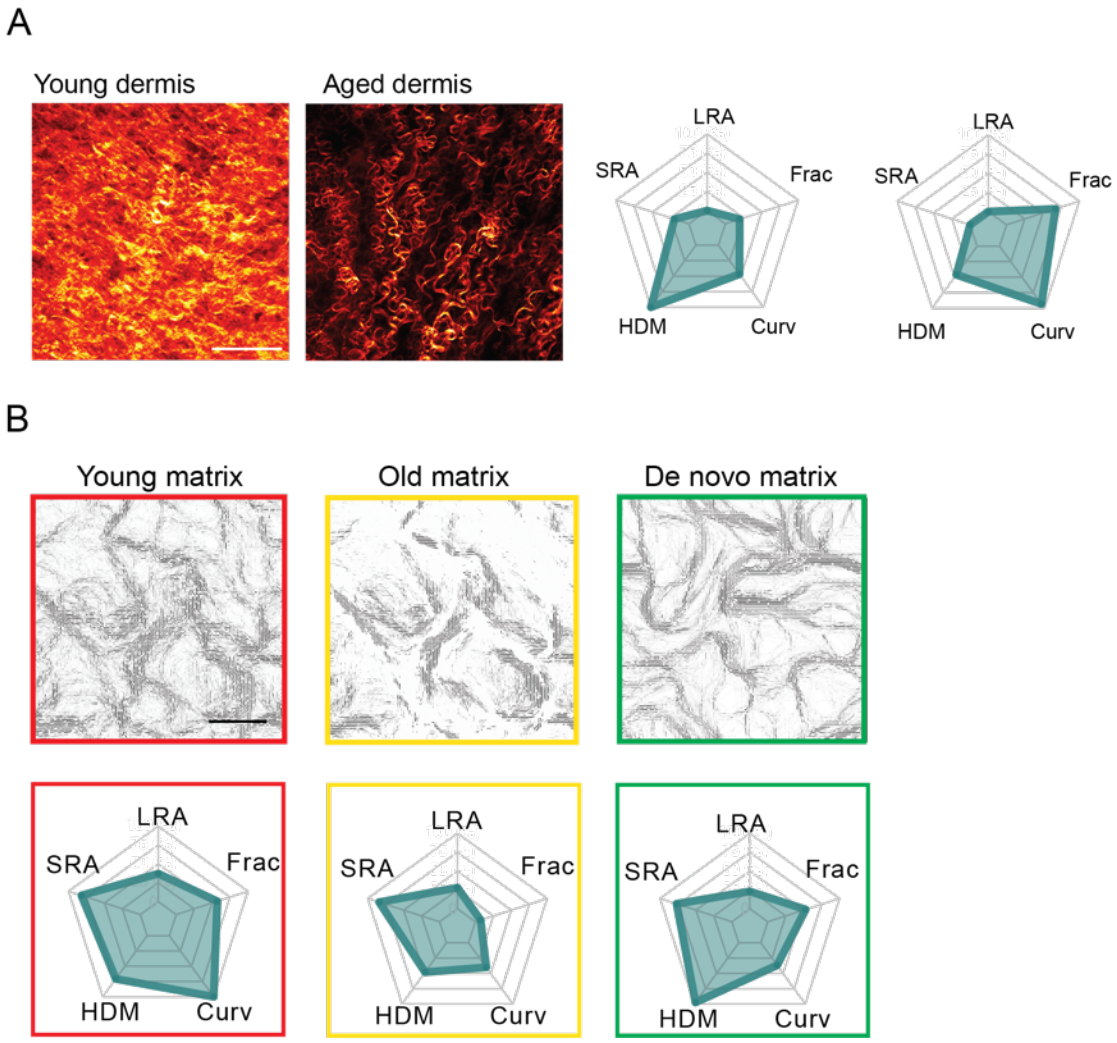


Figure 67: Simulating matrix transitions in ageing

(A) Second harmonic imaging of collagen (orange) in young and old mouse dermis with corresponding starplots to the right. (B) Starting with young matrix, the simulation is run with $w_3 = 0.12, \eta = 0.05, depRate = 1, degRate = 1, reRate = 0$ (middle). The same parameters with the addition of degRate=0 are used to generate a *de novo* matrix (right). Corresponding starplots show how the matrices vary in metric space. Scale bar represents $100\mu m$.

3.8 Chapter highlights

A summary of the main findings of this chapter is given as follows:

- Development of novel, discrete, two-layer Vicsek model to study interplay between fibroblasts and ECM (Wershof *et al.*, 2019)
- A feedback mechanism between fibroblasts and ECM generates diverse patterns reminiscent of *in vivo* tissues.
- Model elucidates how matrix patterns can interconvert over time in cancer, wound healing and ageing.

The relationship between fibroblasts and ECM is a complex one. Fibroblasts are responsible for producing, rearranging and degrading matrix fibres. These fibres provide physical and chemical cues to the environment, thus creating a feedback loop of influence between the fibroblasts and ECM. The model has helped to tease apart the mechanical aspect of this interplay. Future work will need to be carried out addressing the chemical relationship between fibroblasts and ECM. Additionally, it would be useful to incorporate spatial information of cells together with ECM spatial metrics to validate if the cells are behaving similarly to predictions through comparing metrics such as cellular hotspots and colocalization with the matrix. This would build on section 3.6, where the model was challenged to produce matrix patterns that resembled those seen *in vivo* through comparison of matrix metrics alone. The field of matrix conversion is likely to gather interest in the near future, as it has such wide-reaching applications in tissue engineering (Daley, Peters and Larsen, 2008a).

Chapter 4. Quantifying and comparing matrix patterns in patients

4.1 Introduction

In the previous two chapters, the generation of diverse collective fibroblast and matrix patterns was studied. Having explored how matrix patterns are generated, it is interesting to consider what the consequences of such patterning might be. The agent-based model demonstrated the variety of matrix patterns that could be obtained from just small perturbations to the rules of the system, suggesting that this same diverse spectrum of patterns may be observable in cancer. Whilst it has been well documented that aligned matrix is a promoter of cancer cell invasion and can be a negative prognostic marker for survival (Provenzano *et al.*, 2008; Conklin *et al.*, 2011; Drifka *et al.*, 2016b), it is unclear whether other topological matrix properties could be informative. Further, could this paradigm be extended to explore the importance of matrix patterning in other fields, such as developmental biology or fibrotic diseases?

4.2 Data

To answer these questions, a clinical dataset of 105 breast cancer samples was acquired with corresponding clinical data from researchers in Nottingham (Abd El-Rehim *et al.*, 2005; Roxanis *et al.*, 2018). These samples were cut and stained with the following stains, shown in Figure 68:

- Hematoxylin and eosin (H&E) – staining of entire core
- Picosirius red (PSR) – staining of collagen
- Pan cytokeratin (PCK) – staining of cancer cells

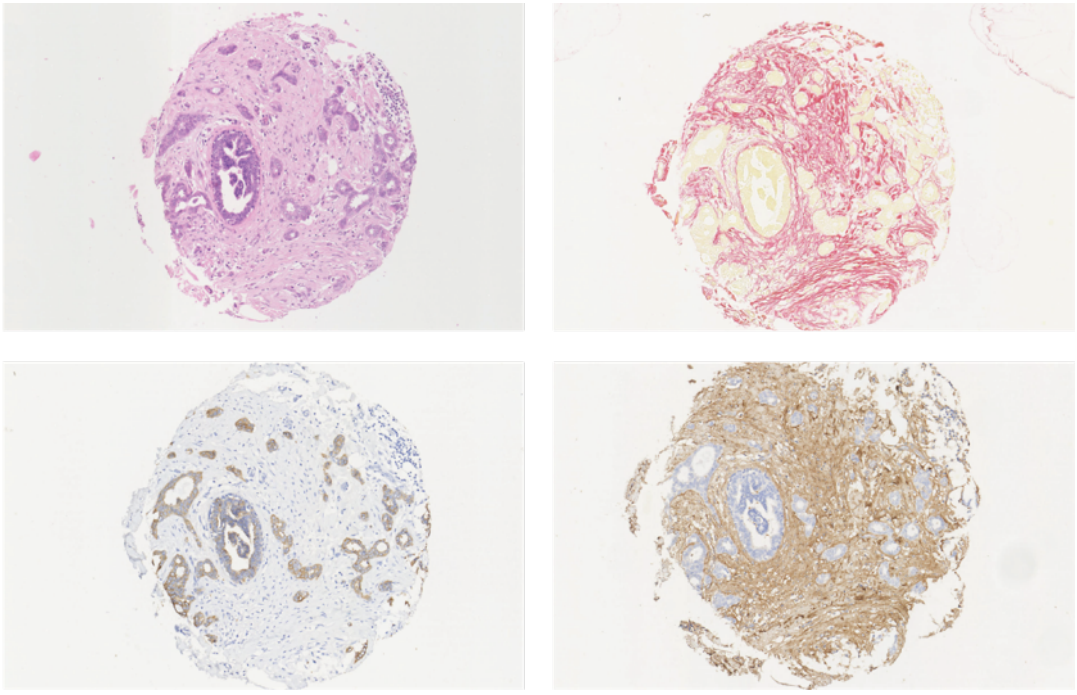


Figure 68: Different staining of breast cancer samples

Top left: H&E staining shows nuclei in blue and cytoplasm/ECM in pink, giving a general indication as to the structure of the sample. Top right: PSR staining in red shows collagen in the sample. Bottom left: PCK staining shows nuclei of cancer cells in blue. Bottom right: composite image of the PCK, PSR with the addition of fibronectin (shown in brown) manually overlaid.

4.3 Metrics for matrix quantification

The metrics introduced in the previous chapter for characterising matrix patterns were extended to consider additional properties of the matrix. Additional metrics were derived from the corresponding derived mask image of the samples (Figure 69). The new metrics included the number of branch points of the matrix and number of end points of the matrix (Figure 70). Number of branch points indicates the number of intersections in the curves of the masks of ECM. Number of endpoints is an intuitive count of the number of ends of the curves in the image. Such metrics are common in analysing spatial information, such as road patterns in urban planning (Reis, Silva and Pinho, 2016). The metrics of branch points and end points were normalised by the total length of the mask curves in order to account for differences in amount of collagen across biopsies. Normalised branchpoints are denoted as NBP, and normalised endpoints is denoted NEP. These extra metrics helped to account for differences in amount of collagen in the samples. Whereas *in*

silico and *in vitro* ECM is laid across the whole sample window, some tumour biopsies are largely made up of adipose or non-matrix tissues. This variability could compromise the robustness of fractal dimension as a metric for complexity. Furthermore, long-range alignment and short-range alignment from the previous chapters were replaced by instead computing the curvature over different curvature windows. Schematics of these metrics, together with examples from the breast cancer dataset are giving in Figure 70.

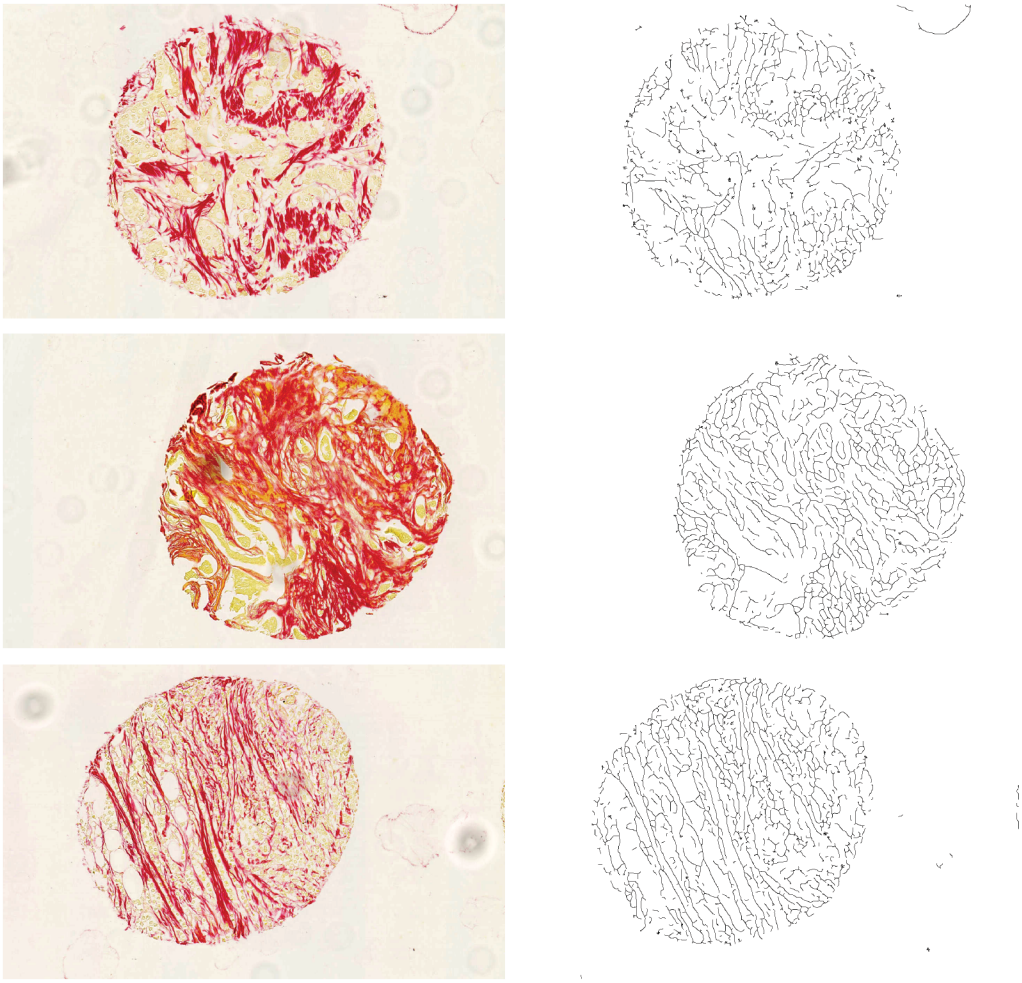


Figure 69: Deriving masks from collagen staining

Masks are derived from tumour cores stained with picosirius red for collagen.

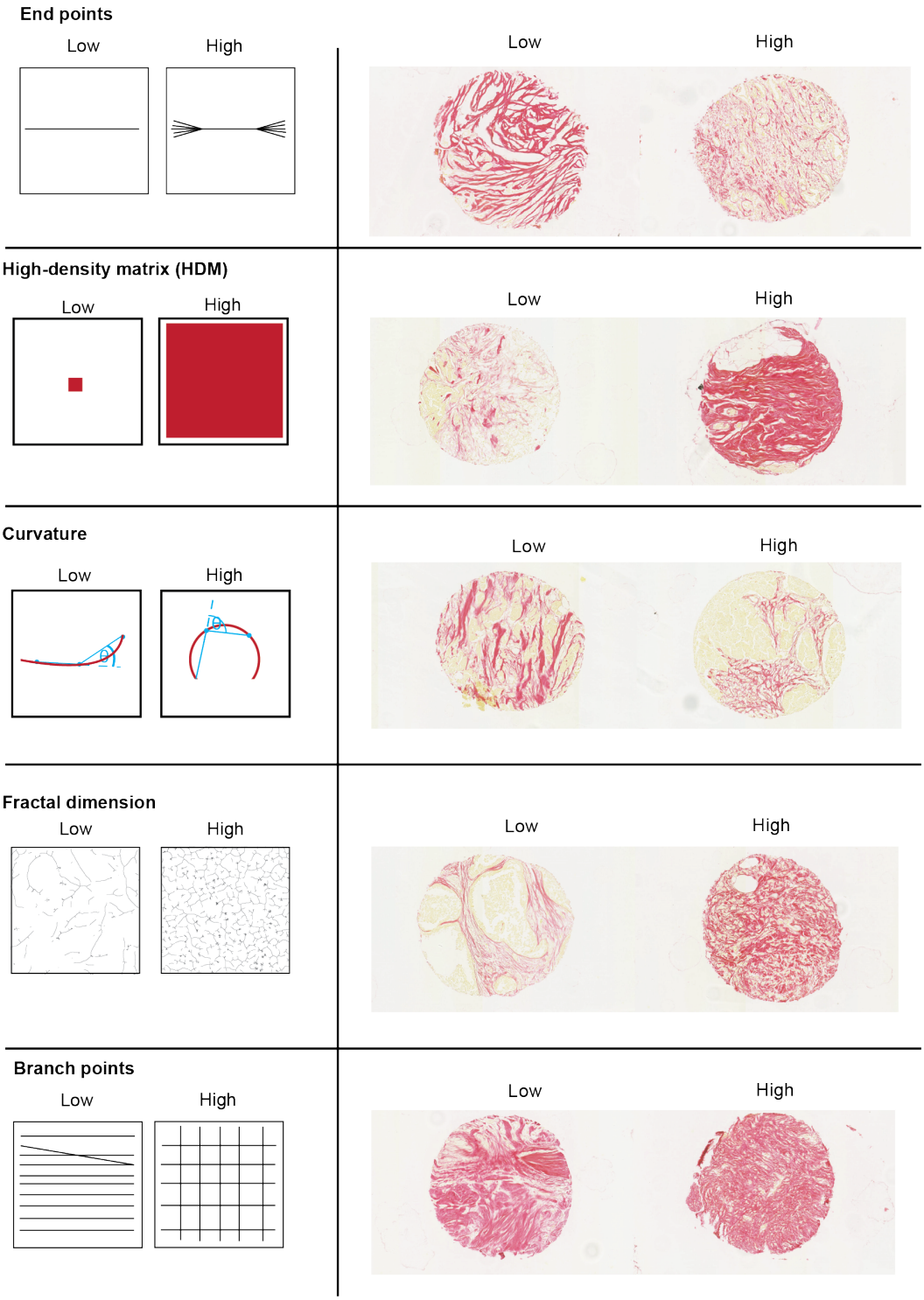


Figure 70: Examples of matrix metrics

Schematics (left) of matrix metrics together with example tumour biopsies (right) showing low and high values of the specified metric. Samples come from the breast cancer dataset described in 4.2. For all metrics, apart from HDM, low and high examples of that metric are shown on samples with comparable levels of HDM.

4.4 Methods

Tumour cores were cut, stained and imaged as described in 6.2.7. Images were then processed in Fiji by converting them to 8-bit images and adjusting the brightness and contrast of the image to optimise clarity of fibres. The plugin Ridge Detection (Steger, 1998) was then used to derive a mask of the ECM. Matrix metrics were then derived using the QuantBlackSpace and Anamorf (Barry, Williams and Chan, 2015) plugins written by David Barry in the Light Microscopy platform at the Francis Crick Institute. Finally, these metrics were combined with the clinical data and Cox survival analysis was performed. A workflow diagram showing this process is given in Figure 71.

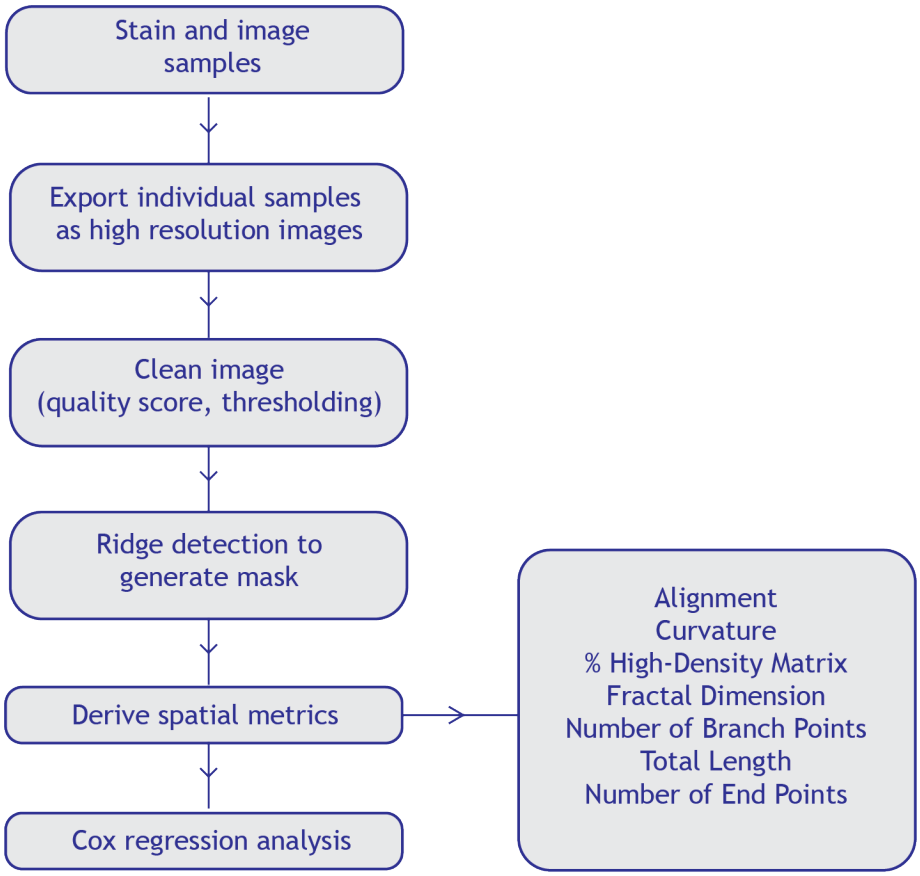


Figure 71: Workflow diagram of quantification of matrix patterns

End-to-end pipeline from obtaining the samples through matrix quantification to survival analysis based on this ECM metrology. A list of metrics is given in the right-most box.

4.5 Results

This methodology was followed to investigate if matrix metrics could be prognostic indicators in the breast cancer dataset from Nottingham. Patients presenting from 1986-1998 with primary operable invasive breast carcinoma were entered into the Nottingham Tenovus Primary Breast Carcinoma Series (Abd El-Rehim *et al.*, 2005; Roxanis *et al.*, 2018). A total of 196 patients were followed until their death or the end of the study, whichever came first. The event (E) of interest is specifically death from breast cancer. The mean age of the patients at age of diagnosis was 53.5 years, with an average tumour size of 2.02mm. A summary of key clinical statistics is given in Table 11. Of these patients, 26 were removed from the analysis due to samples that were damaged or contained too little ECM. These decisions were made in collaboration with Antonio Rullan, a medical oncologist from the Tumour Cell Biology Laboratory at The Francis Crick Institute. A further 65 patients were also not included in the survival analysis due to missing or incomplete clinical data on survival times. This left a total of $N = 105$ patients with complete clinical data and a tumour biopsy with sufficiently high levels of ECM.

Table 11: Summary of breast cancer dataset

Age at diagnosis	mean (53.5)	lower 95% CI (52.2)	upper 95% CI (54.8)	
Tumour size	mean (2.02)	lower 95% CI (1.89)	upper 95% CI (2.16)	
Lymph node status	positive (37)	negative (72)	NA (70)	
Histology subtype	Lobular (26)	Ductal (147)	MixedNST (5)	
Biological subtype	HER2 (17)	ER (72)	triple negative (21)	NA (69)
Nottingham Grade	1 (30)	2 (58)	3 (88)	NA (3)

A Cox proportional hazards regression model was then used to relate matrix metrics to the event of death from breast cancer (E). In addition, tumour size (TS) and tumour grade (TG) were included as model parameters for comparison, as they are well established indicators of survival currently used by clinicians. The analysis was carried out using the “survival” and “survminer” packages in R. The model can then be written as

$$E \sim LRA + HDM + Curv + Frac + NBP + NEP + TS + TG$$

All reasonable combinations of these variables were tried by incrementally removing variables from the model. Only statistically clear results are reported here (with a p-value<0.05).

Unsurprisingly, tumour grade and size are indicators of E. The model is given as:

$$\text{Model 1: } E \sim TS + TG$$

This model has a coefficient of 0.349 for tumour size and 0.408 for tumour grade with a p-value of 0.003 according to a likelihood ratio test.

Of the matrix metrics, normalised branchpoints (NBP) was able to act as a univariate predictor of E. The model is given as:

$$\text{Model 2: } E \sim NBP$$

This model has a coefficient of 0.000779 for NBP with a p-value of 0.01 according to a likelihood ratio test.

The hazard ratio of this model was then computed. The exponential coefficient for NBP was $\exp(0.000779) = 1.00078$, ie for every additional NBP, there is an increased risk of 0.078% of E. To put this into context by scaling, patients had a mean number of 914 NBP with a standard deviation of 422, meaning that the expected increase in hazard relative to an increase of 100 NBP is approximately 8%. To understand this visually, a Kaplan-Meier curve showing the lowest and highest quartiles of the patients according to NBP was plotted (Figure 72, comparing upper and lower quartiles: p-value=0.011).

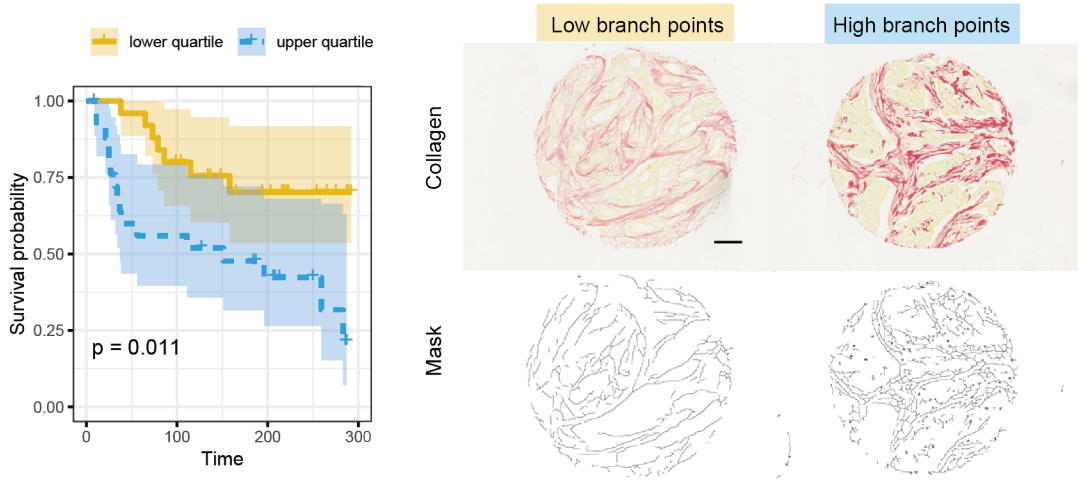


Figure 72: Survival analysis based on matrix metrics.

A Kaplan-Meier curve showing the lowest and highest quartiles of 105 breast cancer patients sorted by number of ECM branch points in their biopsies. (Right) Example slices from biopsies from patients stained for collagen with low (top) and high (bottom) branch points and their respective masks. Time scale is in months. Scale bars represent $\boldsymbol{100\mu m}$.

This information was then combined with the traditional prognostic parameters. The model is given as:

$$\text{Model 3: } E \sim TS + TG + NBP$$

This model has a coefficient of 0.333 for tumour size, 0.437 for tumour grade and 0.000756 for NBP, with a p-value of 0.0005 according to a likelihood ratio test.

Further, models 1 and 3 were compared against each other to see if the additional information of NBP in model 3 produced an improved predictor of E. Indeed, performing ANOVA using the Chi-square test to compare the results showed that model 3 is a statistically significantly better model than just knowing tumour size and grade (anova, p=0.014).

These analyses represent an exciting proof of principle as to the importance of matrix quantification in cancer patient diagnosis. Namely, these results suggest that in breast cancer, the number of branch points in the ECM of tumours is a statistically clear marker of survival. In addition, using this information in conjunction with tumour size and grade provides a better predictor of the event that just tumour size or grade alone.

4.6 Identifying the wider importance of matrix quantification

Whilst these results are promising, the sample size of the breast cancer dataset is rather small. Ideally, other members of the scientific community would be able to conduct similar analysis on their own data sets and validate the power of matrix quantification.

The first challenge of transferring these methods to other datasets is that many tumour samples are stained only with H&E, for example the TCGA dataset. To address this, stromal cells from H&E cores in the breast cancer dataset were cross-correlated with ECM organisation. Khalid Jabbar and Shan Raza from Yinyin Yuan's group at the Institute for Cancer Research performed machine learning analysis on the H&E stained biopsies to identify the position and orientation of fibroblasts. These data were then plotted as vector fields and showed excellent qualitative agreement with the orientation of collagen (Figure 73). In addition, the global alignment of the fibroblasts was computed, together with the global alignment of the collagen mask, by taking the orientation of the tangent at incremental points along the curves of the mask. These alignment scores of the fibroblasts and the collagen were then cross-correlated and show strong agreement (Figure 73, Pearson's product moment correlation, degrees of freedom=220, p-value=3e-13). This suggests that in instances where the matrix is not specifically stained for collagen in biopsies, the position and orientation of stromal cells obtained from H&E images may be sufficient to build a comprehensive picture of matrix organisation.

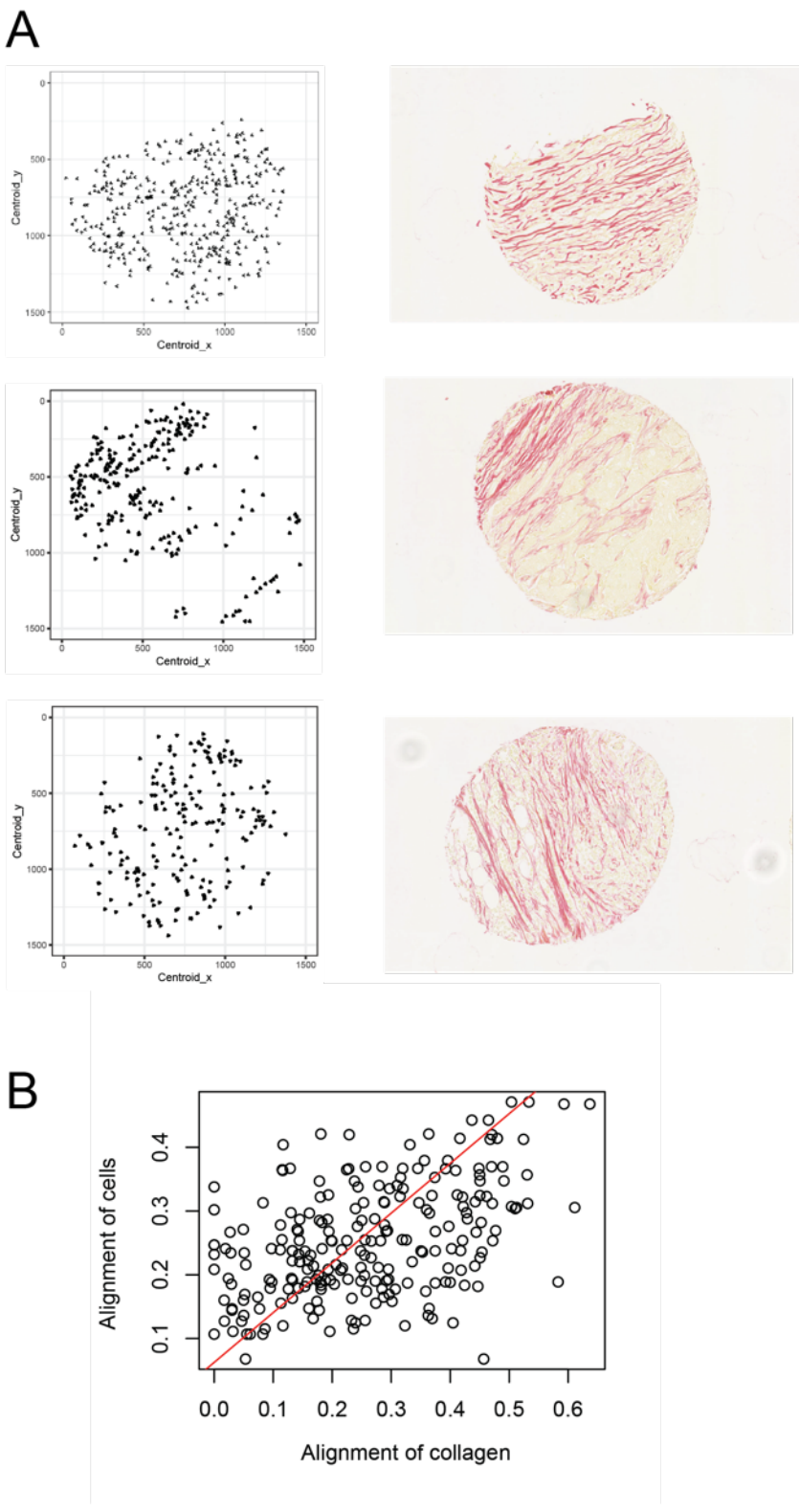


Figure 73: Orientation of fibroblasts correlates with collagen in tumour biopsies

(A) Three example tumour biopsies showing the position and orientation of fibroblasts (left) and collagen staining (right). Qualitatively these are correlated. (B) Comparing global alignment of fibroblasts and collagen fibre orientation shows strong agreement (Pearson's test, p-value<0.0001, gradient=0.46).

The second challenge in applying matrix metrology analysis to other datasets is in the portability and usability of the pipeline by other users to data with different structures and properties. Bioinformatic analysis pipelines often become problem-specific and software-specific, limiting their ability to be repurposed for use on other problems or in different environments (Ewels *et al.*, 2019). To this end, a pipeline is being developed, which will enable robust analysis across datasets. The short-term aim is to generate plugins that can be tested by other members of the ECM community. Longer-term, the pipeline will be refined so that it can be used by the wider community, ranging from experimentalists interested in cell migration and remodelling of 3D matrices, to clinicians researching conditions such as cancer and fibrosis. The generation of quantitative metrics could allow the comparison of results between researchers and the correlation of matrix features with other biological parameters, including transcriptomes and proteomes. Ultimately, the goal is to validate clinical ECM metrics as useful diagnostic or prognostic indicators with a role in clinical decision making. This work is still ongoing, but the progress so far, together with future challenges is reported here.

4.7 Outline of pipeline for deriving matrix metrics (Twombli)

The package developed so far has the working title of Twombli (The Workflow Of Matrix BioLogy), after the American artist whose works are full of diverse marks. Twombli currently consists of a number of ImageJ macros, together with user documentation and a walkthrough example. Twombli consists of three stages:

1) Prechecks – filtering out unsuitable images and appropriately thresholding images
2) Deriving matrix metrics to quantify the matrix patterns
3) Survival analysis on the matrix metrics [optional]

A summary of all the ImageJ macros are given in Table 12.

Table 12: Twombli ImageJ macros

Program/ file name	Function
preChecks.ijm	Interactive user checklist for appropriate file organisation and parameters for thresholding images (Figure 76).
ridgeDetection.ijm	Takes raw collagen images and produces a mask (Figure 77)
runAnamorf.ijm	Derives specified spatial metrics from masks
computeHDM.ijm	Distinguishes foreground (collagen) from background on thresholded images
directionality.ijm	Computes directionality of masks
vectorField.ijm	Takes masks and produces vector field of fibre orientation (optional)
deriveMetrics.ijm	Combines ridgeDetection.ijm, runAnamorf.ijm, computeHDM.ijm, and directionality.ijm macros
SurvivalAnalysis.R	User walkthrough with example survival analysis

4.8 Twombli set up

A user will set up the Twombli master folder, download the macros in a subfolder and arrange their files according to the set up shown in Figure 74. In addition to the macros folder, there are additional empty subfolders for the output masks, output images thresholded for the HDM metric, and a results folder. There is additionally a subfolder in which the user puts the original files. Some of these original files are copied into the TestSet folder. After running the preChecks.ijm macro, these original image files are split into eligible and ineligible images. The rest of Twombli is then run on the eligible images only.

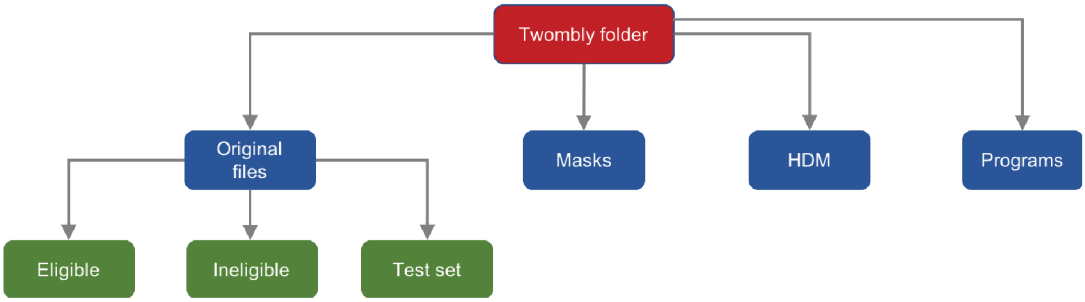


Figure 74: Twombli file organisation

Documentation explains to the user how to arrange files. This includes a folder of Test set images. Images are then subdivided into Eligible and Ineligible folders. The output of the pipeline creates masks of the original images and copies of the images that have been thresholded for HDM.

The aims of the preChecks stage are two-fold. Firstly, ineligible images need to be excluded from the analysis. Images could be unsuitable if the sample has been damaged, if there is too much adipose tissue (and therefore not enough ECM), or if the image resolution is too poor (Figure 75). This division into eligible and ineligible files is currently done manually by the user, with an example in the Twombli walkthrough section. Secondly, eligible images must be suitably adjusted and thresholded in order to optimise the quality of the corresponding image mask. This includes choosing a threshold value for HDM, adjusting the image brightness and contrast and choosing the line width for the ridge detection algorithm. This thresholding is done in the preChecks.im macro where a simple step-by-step user interface helps users choose appropriate values for the images in the testSet folder (Figure 76).

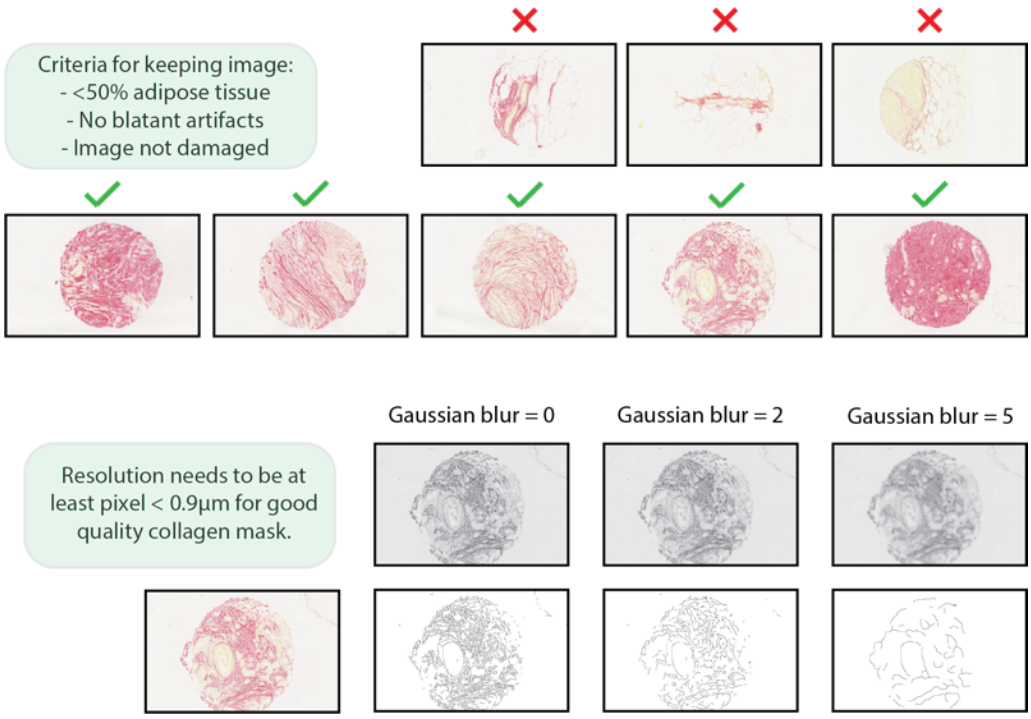


Figure 75: Removing ineligible samples

User walkthrough showing examples images that are ineligible for being poor quality or low resolution.

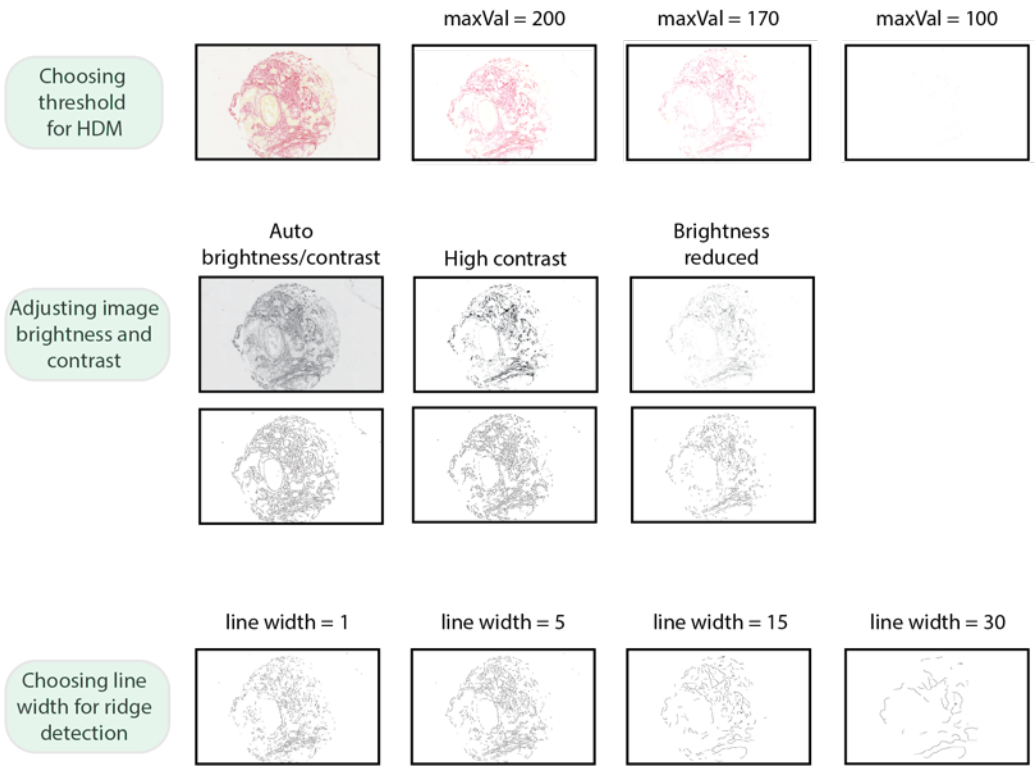


Figure 76: Choosing threshold values for image processing

User walkthrough showing examples of how to threshold images.

4.9 Deriving matrix metrics and survival analysis with Twombli

Having completed the preChecks stage of the process, deriving the spatial metrics is reasonably straightforward. The deriveMetrics.ijm macro takes the outputs from the preChecks thresholding as inputs and applies them to the eligible images folder. The macro runs algorithms to compute a mask image for each eligible image and then derives the quantification of the mask. The output of this process are masks in the Mask subfolder (Figure 77).

Finally, the user walkthrough gives an example of how these matrix metrics can be combined with clinical data to perform survival analysis, potentially identifying relationships between matrix metrology and patient outcome. This stage is however optional, as the aim is that Twombli can be used for a broader scope of pattern recognition in many different normal and pathological tissues than just identifying prognostic markers.

4.10 Development of the pipeline

Features are being added incrementally to Twombli. At each stage the pipeline is beta-tested to identify bugs and difficulties in use. Further, the code is being checked and improved by David Barry in the Light Microscopy Team and Robert Jenkins in the Tumour Cell Biology laboratory at the Francis Crick Institute. A currently private Github repository is being used to develop and check the code as new features are added.

4.11 Challenges for Twombli

The primary challenges in making Twombli a feasible tool for other scientists is in making the results robust. Currently, the user manually decides which images are eligible and ineligible for matrix quantification. Ideally, Twombli will be extended to automate this process, producing an image quality score for eligibility. Similarly, the thresholding values chosen in the preChecks.ijm macro are also user-defined. Analysis needs to be carried out as to how robust results are to differences in user-chosen threshold values, or ideally thresholding could be automated taking away

all manual components of the pipeline. This represents a significant problem which will be tackled in the future (Ewels *et al.*, 2019).

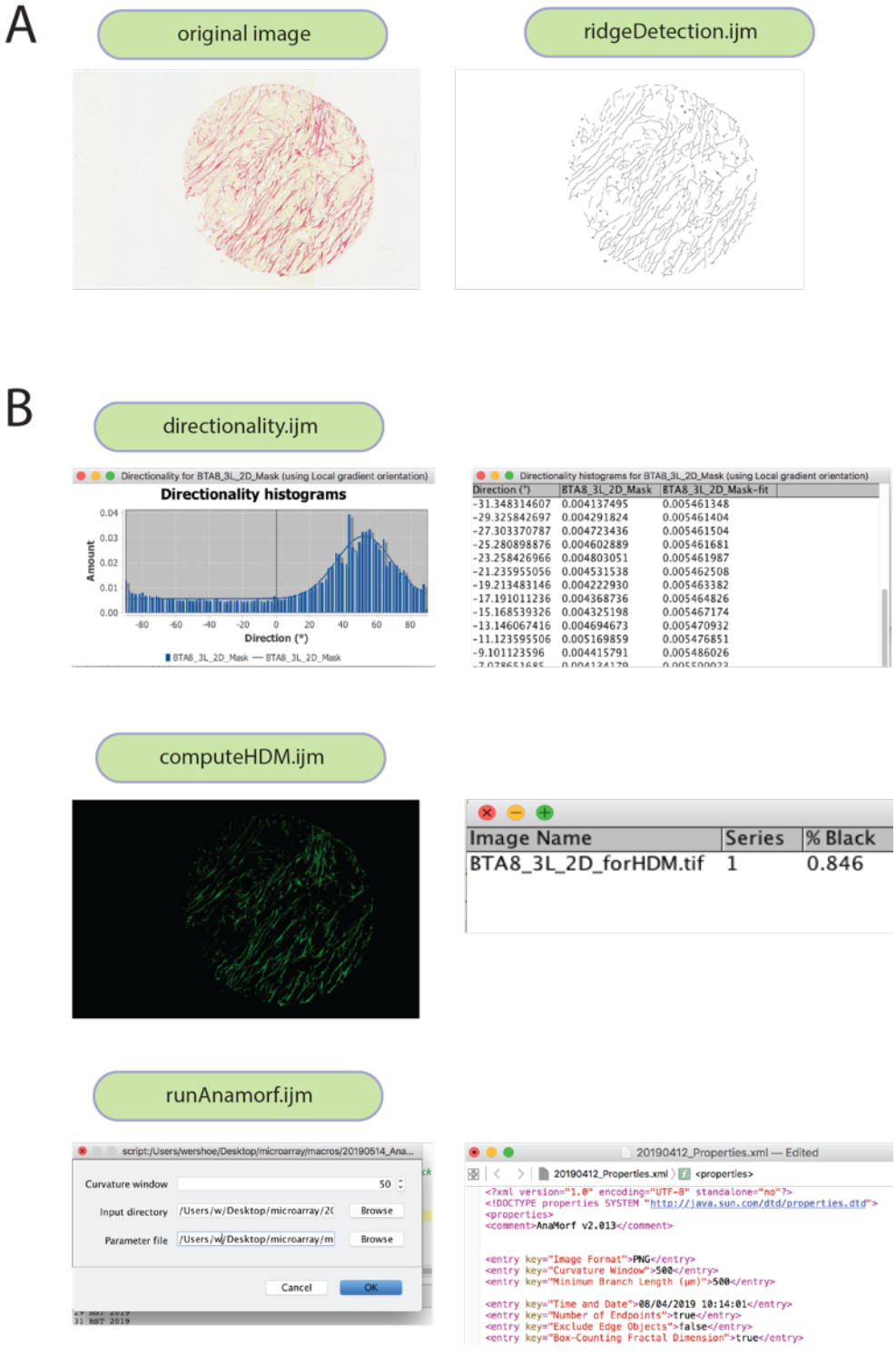


Figure 77: Twombli user interface

(A) Example input image with the output mask from the ridgeDetection.ijm macro. (B) Example of Twombli interface with different macros.

Additionally, once these issues have been addressed and Twombli has been beta-tested by other scientists, the long-term aim will be to adapt Twombli to a user

interface that could be used by clinicians and pathologists. This will most likely require sophisticated software development, but represents a promising future direction of work with the potential to have meaningful impact in the clinical environment.

4.12 Future application of pipeline

In conjunction with the future developments to Twombli as described above, the importance of matrix metrics needs to be tested on a larger dataset. Additionally, it is of interest to see how different matrix patterns could be indicative of different tumour properties in different kinds of cancers. To this end, similar analysis will be carried out on a dataset of 2000 prostate cancer patients from The Royal Marsden Hospital (A. Wilkins *et al.*, 2018). To date, this biobank has been used entirely to explore tumour gland biology (A. C. Wilkins *et al.*, 2018). No work has yet been carried out looking at the tumour stroma.

The novelty of Twombli lies in both its versatility and universality. Twombli can be applied to diverse structures collected via many different methods, for example mycelium (Obert, Pfeifer and Sernetz, 1990; Barry, Williams and Chan, 2015) or extracellular matrix images collected from freeze drying microscopy (Figure 78). Other work in the field of matrix quantification has looked principally at alignment (Rezakhaniha *et al.*, 2012; Park *et al.*, 2019) and often require heavy manual intervention. Twombli not only unifies many different matrix metrics for the first time, but also largely automates an end-to-end pipeline whilst providing guidelines on how to appropriately parameterise and filter images for robust results.

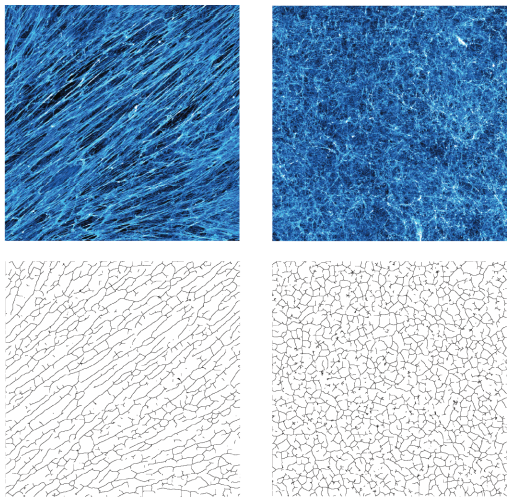


Figure 78: Twombli works on versatile images.

ECM patterns obtained *in vitro* through freeze drying microscopy methods. Twombli captures underlying structure of ECM.

4.13 Chapter highlights

A summary of the main findings of this chapter is given as follows:

- Matrix patterns with a high number of branch points are predictive of poor survival outcomes in 105 breast cancer patients.
- An end-to-end pipeline called Twombli has been developed for quantification of matrix patterns.
- A plan has been developed to extend Twombli to account for robustness of results, automate image thresholding and improve user-interface for potential use in clinical decision-making.

In conclusion, it is clear that quantification of the ECM represents a promising and as yet largely unexplored route for predicting tumour spatial evolution. Whilst work has established that alignment of collagen is a prognostic marker for survival (Conklin *et al.*, 2011; Drifka *et al.*, 2016a; Best *et al.*, 2019), the results presented here suggest that matrix organization can be far more nuanced. The complex architecture of the ECM may provide clues as to how tumours are likely to develop, and this work will be continued in the near future.

Chapter 5. Discussion

5.1 Summary of results

In the introduction, three research questions were posed. Below are answers to these questions based on the work presented in this thesis:

Q1 What properties of CAFs cause different collective behaviours?

In Chapter 2, it was shown that cell persistence level is a key contributing factor that determines overall alignment. If cells have poor persistence, no reasonable level of collision guidance is able to cause global alignment. The results suggest that the non-aligning CAFs fall into this category. However, high persistence alone is not sufficient for generating alignment and requires an active mechanism of cell-cell collision guidance. Cell-cell collision guidance can change the speed at which alignment occurs and for moderate levels of cell persistence, becomes critical in determining the emergent pattern. Experimentally, the aligning CAFs display a mid-range persistence and therefore the model suggests a moderate level of collision guidance must necessarily be at play in order to facilitate alignment. The model indicates that for long-range alignment to occur, a set of multi-scale conditions must be satisfied: At the level of the individual cell, cells must migrate with sufficiently high persistence. At the collective level, cells must influence the orientation of each other and be of a sufficiently high density.

An understanding of the contribution of these parameters allows for exploration of alignment in heterogeneous populations. The model predicted the emergent alignment behaviour in a heterogeneous CAF population and illustrated how the cell motility properties of each population determines the proportions of each required to disrupt alignment. Analyses indicate that there are certain instances where a low proportion of non-aligning CAF would be sufficient to disrupt alignment. This offers a mechanistic view on how limited alignment in a highly heterogeneous tumour micro-environment should be relative easy to obtain. Given the fragility of emergent alignment, for an area of alignment to occur, the population would need to be largely dominated by CAFs displaying an aligning phenotype. If

this sensitivity of alignment applies *in vivo,* only small phenotypic perturbations would be required to dissolve harmful microstates of alignment. Whilst much literature has focussed on analysing cancer cell heterogeneity (McGranahan and Swanton, 2017), there has been a lack of both experimental and computational work examining fibroblast heterogeneity and their collective behaviour. The model demonstrates this to be a promising future area of research.

Q2 What is the effect of the mechanistic interplay between CAFs and ECM?

Fibroblasts and ECM have a fundamental dependence on each other (Levental *et al.*, 2009; Frantz, Stewart and Weaver, 2010; van Helvert, Storm and Friedl, 2018). Whilst the effects of ECM geometry on a single cell have been documented (Tozluoğlu *et al.*, 2013b; Kim *et al.*, 2015), the consequences of these interactions at the mesoscale have been largely unexplored. The model from Chapter 2 was extended to include a second layer, consisting of a grid of matrix fibres. Therefore, the model comprised two layers: a top layer of fibroblasts, and an underlying layer of extracellular matrix. The main innovation to the original Vicsek model was in adding the ability of cells to be guided by the matrix fibres below it whilst simultaneously producing and reorganise these fibres. This feedback causes the matrix fibres to become more oriented like the fibroblasts and the fibroblasts to become more oriented to the matrix fibres causing a “flocking-like” behaviour between fibroblasts and matrix. Whilst the matrix fibres do not have a forward motility, the averaging that takes place between the fibroblasts and the matrix reminded us of flocking behaviour.

Such a model was able to shed light on the co-evolution of fibroblasts and ECM. Notably, the introduction of this feedback mechanism was responsible for generating far more diverse patterns of ECM, some of which were reminiscent of *in vivo* tissues. In particular, matrix feedback gave rise to patterns with a wide range of curvatures. It is currently unclear how many ECM patterns *in vivo* arise and the model suggests that matrix feedback is a mechanism that could possibly facilitate this emergent patterning. Additionally, the model was used to explore how matrix patterns could transition over time as a consequence of cancer, ageing or wound healing. There is far more work to be done in this area, with experiments

specifically looking at fibre reorganisation in order to choose better model parameters. However, the model is ideally suited to exploring these matrix conversion questions in more detail and agent-based modelling in general is well placed to observe spatial and temporal dynamics together (Dallon, Sherratt and Maini, 1999), which remains a big challenge experimentally.

Q3 Can spatial organisation of matrix predict patient outcome?

Whilst the first two questions dealt with how emergent matrix patterns are generated, it was of great interest to understand the clinical consequences of such patterns. To this end, collagen-stained tumour biopsies from breast cancer patients were analysed. Quantification of matrix metrics revealed that the number of matrix branch points was a predictor of survival above and beyond knowing just tumour grade and size. Spurred on by this finding, an end-to-end pipeline called Twombli was established and is still in development. The goal is to use Twombli to analyse other data sets of matrix patterns, for further evidence of the value of matrix quantification and ultimately, to enable clinicians to use the software in clinical decision-making. A clear caveat to this approach is in how the algorithm will deal with other tumour components. However, with the increasing ease of tumour imaging it is likely that matrix and tumour architecture will become more important as a prognostic tool in the near future (Nawaz *et al.*, 2015; Yuan, 2016).

These three questions all relate to aspects of TME geography. Understanding how patterns in tumours form and are likely to develop remains a key target for cancer therapy (Hirata and Sahai, 2017). A complementarity between mathematical modelling and experimentation has helped to elucidate this evolution. Furthermore, the model has helped to ask important questions, some of which remain unanswered:

- Could cell-cell collision geometry have an impact on cellular response i.e. head-head collision vs head-body collisions (Desai *et al.*, 2013)?
- Are fibroblasts more inclined to follow newer/fresher fibres?
- How far-reaching are MMPs secreted from fibroblasts i.e. long-range degradation?

5.2 If time travel existed…

5.2.1 Going back in time

With the benefit of hindsight, there are some aspects of the project that ought to have been done differently. The most substantial difficulty was in accurately parameterising the model. Experiments were often carried out under slightly different conditions, with a microscope set at a different scale and frame rate across experiments, making parameter fitting prone to errors. Additionally, a parameter might be fitted according to one experiment, and then be different in another experiment where another parameter was being measured. This made acquiring parameters tricky. Ultimately, this did not have a profound effect on the model, which was used to explore the entirety of parameter space, but was problematic in exactly fitting the simulations to experiments. Given the chance to do it again, it would have been wise to discuss and plan for these difficulties before the experiments had been carried out. In addition, when evaluating the individual migratory noise of cells, it is clear that sometimes the cells undergo occasional random repolarisation (as seen by comparing Figure 20 with Figure 21), something which is not considered in the model. This phenomenon has been previously documented (Desai *et al.*, 2013) and it would be useful to gauge the effect of this repolarisation on the collective behaviour of the system.

It would also have been insightful to have both computationally and experimentally explored additional matrix properties such as crosslinking and matrix stiffness (Paluch *et al.*, 2015; van Helvert, Storm and Friedl, 2018). The model could have been used to explore what the effect of different matrix stiffness values would be on emergent patterning. Multiple matrix components, such as different collagens and fibronectin could also have been incorporated for their different properties. Considering just one type of matrix is likely an oversimplification of the problem, particularly within the context of wound healing where fibronectin plays a crucial role (Dallon, Sherratt and Maini, 1999; Cox and Erler, 2011).

Finally, the project would have benefitted from looking at clinical data earlier on. However, as described by Uri Alon (Alon, 2009), research nearly always throws

unexpected results. Whilst starting at point A and trying to get to point B, one will almost invariably end up instead at point C *via* a highly nonlinear route. This unpredictability though is what makes science so exciting.

5.2.2 Going forward in time

One area of research that has not been looked at in this thesis is the phase transition space between disorder, spatially homogeneous order and spatially heterogeneous order of the model. It has been shown in (Escaff *et al.*, 2018) that varying interaction range and interaction strength between particles of a similar continuous system can generate all three behaviours. Work by (Chaté *et al.*, 2008; Duclos *et al.*, 2014) has studied more general properties of flocking systems such as giant-number fluctuations and correlation length of emerging order. It would be of great interest to apply similar levels of physical analysis to the model presented in this work, in particular to see the effects of matrix feedback on the emergence of order and phase transitions between different states.

Similarly, the model could be used to explore the effects of matrix feedback on jamming (Sadati *et al.*, 2013). With stronger and more carefully constructed volume exclusion rules to circumvent the constant speed of cells in a traditional Vicsek model (Vicsek *et al.*, 1995), the model would be better able to model high-confluence systems in which jamming can occur. Traditionally Voronoi/Vertex models are employed for such systems (Camley and Rappel, 2017), however, the model correctly adapted could be highly informative in what is a largely unintuitive problem: does matrix feedback help or hinder jamming/unjamming?

In the agent-based model presented here, long-range chemical signalling between fibroblasts is not considered. This is an avenue of future research that should be addressed. Vicsek-like models have explored zones of repulsion, alignment and long-range attractions for general particles (Vicsek and Zafiris, 2010) and in the collective migration of ducks (Lukeman, Li and Edelstein-Keshet, 2010). By tracking the movement of individual ducks in the flock, Lukeman et al. are able to infer that ducks move with zones of repulsion, alignment and attraction with their neighbours, with an additional attraction/repulsion interaction with a single

neighbour in front due to easier visual awareness of that neighbour. There is no evidence of leaders and followers in such a group. It remains to be seen if a similar preference for particular neighbours could be observed in cell flocking behaviour. Such analysis would require more detailed cell collision analysis, perhaps through constructing matrix micropatterns that cells would be induced to follow, allowing for the observation of specific and planned collisions. Additionally, the model could be extended to three dimensions (Schumacher, Maini and Baker, 2017), coupled with carefully constructed three-dimensional experiments facilitating cell-cell collisions.

An essential consideration in understanding the spatial architecture of the tumour is the nature of heterotypic interactions between different cell types (Valastyan and Weinberg, 2011). Recent work has shown that it is heterotypic E-cadherin/N-cadherin junctions between CAFs and cancer cells which allow CAFs to exert a pulling force on the cancer cells, thereby promoting invasion (Labernadie *et al.*, 2017). Furthermore, it has been shown in several cancers that cancer cells lose heterotypic CIL response whilst maintaining homotypic CIL, leading to cancer cell invasion of other tissues (Mayor and Carmona-Fontaine, 2010; Roycroft and Mayor, 2015; Stramer and Mayor, 2016). It would of interest to couple an understanding of heterotypic interactions with the spatial map of the tumour margin in order to predict points of metastasis. This would also link back to the idea of Tumour Associated Collagen Signatures (TACS), which consider collagen arrangement with respect to the tumour mass (Provenzano *et al.*, 2006).

Following on from this, it would be exciting to extend Twombli to incorporate the generation of a TACS score. In order to facilitate this, tissue samples of ECM at the tumour boundary would need to be analysed. Obtaining such samples would be a significant challenge. Furthermore, Twombli could also incorporate information on topological defects (Hirst and Charras, 2014; Saw *et al.*, 2017). Whilst topological defects have not been considered in this work, they are an emerging area of interest in cancer invasion (Saw *et al.*, 2017) and it would be useful to include this additional spatial information in matrix pattern quantification.

Complementary to this, deep learning methods could be hugely useful in identifying more abstract patterns in patient biopsies. The model thus far has been used to

understand how different cellular mechanisms give rise to different collective alignment behaviours. Starting with given initial conditions, the model can predict what the emergent cell patterning will be at the end. However, in a clinical setting it is rarely the case that the information presented starts at the beginning. Rather, emergent patterns have already formed. Could the model instead be used to work backwards? It would be of great help if, given a matrix pattern (and by extension, pattern of fibroblast trajectories), one could infer the properties of the cells that generated such an ECM. This would be useful for two reasons: firstly, knowing the cellular makeup of a tumour could inform therapeutic decision-making. Secondly, knowing the mechanistic properties of cells in a tumour would give an indication as to how the tumour structure was likely to evolve both spatially and temporally. Such a reverse-engineering challenge would required the complementary use of the agent-based model and machine learning (Baker *et al.*, 2018) This work is ongoing and is being carried out in collaboration with Xiao Fu, a postdoc in the Biomolecular Modelling Laboratory, who is developing a CNN trained on synthetic data generated by the agent-based model. Deep learning is a rapidly advancing field and it is clear that its coupling with traditional mathematical modelling techniques will massively further our understanding of tumour evolution.

Finally, it is important to consider the implications of this work on cancer therapies. The tumour microenvironment plays a key role in cancer progression and much effort is being put into treatments that target the TME in conjunction with, or as an alternative to, traditional treatments targeting cancer cells (Hirata and Sahai, 2017). So far, the results of these approaches have not met expectations. One such example is in targeting CAFs directly. The disappointing results are possibly due to CAFs having both pro-tumour and anti-tumour effects (Özdemir *et al.*, 2014; LeBleu and Kalluri, 2018). Another example is of therapies to reduce the tumour vasculature. However, a side effect of this approach is a hypoxic environment, favouring cancer cells with activated HIF1α, which has been associated with increased tumour invasion (El-Naggar *et al.*, 2015; Hirata and Sahai, 2017).

This thesis elucidates the diversity of spatial patterning of CAFs and ECM. Therapies focused on altering CAF/ECM patterning have been largely unexplored. Introducing a small population of non-aligning fibroblasts could help remodel

aligned matrices thereby hindering invasion (2.7.7) (Drifka *et al.*, 2016b). Alternatively, increasing CAF-ECM feedback could result in the generation of dense swirl-like matrices that are difficult for cancer cells to penetrate, whilst still being porous enough to allow immune cell infiltration. Lastly, therapies directly targeting the ECM have the potential to dramatically alter the architecture of a tumour towards an invasion-inhibiting structure and are beginning to be developed with promising results in mice (Natarajan *et al.*, 2019). Twombli could possibly aid in identifying exactly what such invasion-inhibiting structure might look like.

5.3 Final thoughts

Coupling cell-cell and cell-matrix interactions together, the model has elucidated how a minimal set of rules can produce diverse tissue patterns. The multi-layered flocking model offers a potential hypothesis for how many of the complex patterns seen *in vivo* could arise at the mesoscale level. Unifying cell and matrix behaviours has helped to provide new mechanistic understanding of the consequences of matrix feedback, including its importance in the generation of curved matrix patterns. In the field of matrix remodelling, predictions from the model indicate that high levels of feedback hamper the interconversion between matrix patterns. These findings are likely to be of use in the emerging field of tissue engineering (Daley, Peters and Larsen, 2008b).

It is clear that in the TME there are multiple promising lines of attack to disrupt CAF alignment and other harmful patterns. The flocking model developed here provides a simple set of rules to describe mechanisms involved in generating higher-order patterning. As tumour mapping technology becomes more widespread (Heindl, Nawaz and Yuan, 2015; Leung, Rice and Barton, 2015), with concomitant increases in levels of high quality image data, and in conjunction with mathematical modelling and machine learning, it will become possible to predict where certain patterns will occur and their consequences for tumour invasion.

Reference List

Abd El-Rehim, D. M. *et al.* (2005) 'High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses', *International Journal of Cancer*. doi: 10.1002/ijc.21004.

Ahmadzadeh, H. *et al.* (2017) 'Modeling the two-way feedback between contractility and matrix realignment reveals a nonlinear mode of cancer cell invasion', *Proceedings of the National Academy of Sciences*, p. 201617037. doi: 10.1073/pnas.1617037114.

Albert, P. J. and Schwarz, U. S. (2016) 'Modeling cell shape and dynamics on micropatterns', *Cell Adhesion and Migration*, 10(5), pp. 516–528. doi: 10.1080/19336918.2016.1148864.

Alon, U. (2009) 'How To Choose a Good Scientific Problem', *Molecular Cell*, 35(6), pp. 726–728. doi: 10.1016/j.molcel.2009.09.013.

Ananthakrishnan, R. and Ehrlicher, A. (2007) 'The forces behind cell movement', *International Journal of Biological Sciences*, 3(5), pp. 303–317. doi: 10.7150/ijbs.3.303.

Anderson, A. R. A. *et al.* (2007) 'Mathematical Modelling of Tumour Invasion and Metastasis', *Journal of Theoretical Medicine*. doi: 10.1080/10273660008833042.

Augsten, M. (2014) 'Cancer-associated fibroblasts as another polarized cell type of the tumor microenvironment.', *Frontiers in oncology*, 4(March), p. 62. doi: 10.3389/fonc.2014.00062.

Augsten, M. *et al.* (2014) 'Emergent behaviors from a cellular automaton model for invasive tumor growth in heterogeneous microenvironments.', *PLOS Computational Biology*, 11(3), pp. 175–192. doi: 10.3389/fonc.2014.00062.

Baglietto, G., Albano, E. V. and Candia, J. (2012) 'Criticality and the onset of ordering in the standard Vicsek model', *Interface Focus*. doi: 10.1098/rsfs.2012.0021.

Baker, R. E. *et al.* (2018) 'Mechanistic models versus machine learning, a fight worth fighting for the biological community?', *Biology Letters*. doi:

10.1098/rsbl.2017.0660.
Baneyx, G. and Vogel, V. (1999) ‘Self-assembly of fibronectin into fibrillar networks underneath dipalmitoyl phosphatidylcholine monolayers: Role of lipid matrix and tensile forces’, *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.96.22.12518.
Barry, D. J., Williams, G. A. and Chan, C. (2015) ‘Automated analysis of filamentous microbial morphology with AnaMorf’, *Biotechnology Progress*, 31(3), pp. 849–852. doi: 10.1002/btpr.2087.
Basan, M. *et al.* (2013) ‘Alignment of cellular motility forces with tissue flow as a mechanism for efficient wound healing’, *Proceedings of the National Academy of Sciences of the United States of America*, 110(7), pp. 2452–9. doi: 10.1073/pnas.1219937110.
Battersby, S. (2015) ‘News Feature: The cells that flock together’, *Proceedings of the National Academy of Sciences*, 112(26), pp. 7883–7885. doi: 10.1073/pnas.1508834112.
Bearer, E. L. *et al.* (2009) ‘Multiparameter computational modeling of tumor invasion’, *Cancer Research*, 69(10), pp. 4493–4501. doi: 10.1158/0008-5472.CAN-08-3834.
Bentley, K. *et al.* (2014) ‘The role of differential VE-cadherin dynamics in cell rearrangement during angiogenesis’, *Nature Cell Biology*, 16(4), pp. 309–321. doi: 10.1038/ncb2926.
Bentley, K., Gerhardt, H. and Bates, P. A. (2008) ‘Agent-based simulation of notch-mediated tip cell selection in angiogenic sprout initialisation’, *Journal of Theoretical Biology*, 250(1), pp. 25–36. doi: 10.1016/j.jtbi.2007.09.015.
Best, S. L. *et al.* (2019) ‘Collagen organization of renal cell carcinoma differs between low and high grade tumors’, *BMC Cancer*, 19(1), p. 490. doi: 10.1186/s12885-019-5708-z.
Bredfeldt, J. S. *et al.* (2014) ‘Computational segmentation of collagen fibers from second-harmonic generation images of breast cancer’, *Journal of Biomedical Optics*. doi: 10.1117/1.JBO.19.1.016007.
Bulten, W. *et al.* (2019) ‘Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard’,

Scientific Reports. doi: 10.1038/s41598-018-37257-4.
Calvo, F. *et al.* (2013) 'Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts', *Nat Cell Biol*. Nature Publishing Group, 15(6), pp. 637–646. doi: 10.1038/ncb2756.
Camley, B. A. *et al.* (2014) 'Polarity mechanisms such as contact inhibition of locomotion regulate persistent rotational motion of mammalian cells on micropatterns', *Proceedings of the National Academy of Sciences*, 111(41), pp. 14770–14775. doi: 10.1073/pnas.1414498111.
Camley, B. A. *et al.* (2016) 'Emergent Collective Chemotaxis without Single-Cell Gradient Sensing', *Physical Review Letters*, 116(9). doi: 10.1103/PhysRevLett.116.098101.
Camley, B. and Rappel, W.-J. (2017) 'Physical models of collective cell motility: from cell to tissue', *Journal of Physics D: Applied Physics*, 50, p. 113002. doi: 10.1088/1361-6463/aa56fe.
Carmona-Fontaine, C. *et al.* (2017) 'Metabolic origins of spatial organization in the tumor microenvironment', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 114(11), pp. 2934–2939. doi: 10.1073/pnas.1700600114.
Chang, S. S. *et al.* (2013) 'Guidance of cell migration by substrate dimension', *Biophysical Journal*, 104(2), pp. 313–321. doi: 10.1016/j.bpj.2012.12.001.
Chaté, H. *et al.* (2008) 'Modeling collective motion: Variations on the Vicsek model', *European Physical Journal B*, 64(3–4), pp. 451–456. doi: 10.1140/epjb/e2008-00275-9.
Conklin, M. W. *et al.* (2011) 'Aligned collagen is a prognostic signature for survival in human breast carcinoma', *American Journal of Pathology*, 178(3), pp. 1221–1232. doi: 10.1016/j.ajpath.2010.11.076.
Couzin, I. D. *et al.* (2002) 'Collective memory and spatial sorting in animal groups', *Journal of Theoretical Biology*. doi: 10.1006/jtbi.2002.3065.
Cox, T. R. and Erler, J. T. (2011) 'Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer', *Disease Models & Mechanisms*, 4(2), pp. 165–178. doi: 10.1242/dmm.004077.

Daley, W. P., Peters, S. B. and Larsen, M. (2008a) ‘Extracellular matrix dynamics in development and regenerative medicine’, *Journal of Cell Science*, 121(3), pp. 255 LP – 264. Available at: http://jcs.biologists.org/content/121/3/255.abstract.
Daley, W. P., Peters, S. B. and Larsen, M. (2008b) ‘Extracellular matrix dynamics in development and regenerative medicine’. doi: 10.1242/jcs.006064.
Dallon, J. C., Sherratt, J. A. and Maini, P. K. (1999) ‘Mathematical modelling of extracellular matrix dynamics using discrete cells: Fiber orientation and tissue regeneration’, *Journal of Theoretical Biology*. doi: 10.1006/jtbi.1999.0971.
Davis, J. R. *et al.* (2015) ‘Inter-cellular forces orchestrate contact inhibition of locomotion’, *Cell*. The Authors, 161(2), pp. 361–373. doi: 10.1016/j.cell.2015.02.015.
Desai, R. A. *et al.* (2013) ‘Contact inhibition of locomotion probabilities drive solitary versus collective cell migration’, *Journal of The Royal Society Interface*, 10(88), pp. 20130717–20130717. doi: 10.1098/rsif.2013.0717.
Drifka, C. R. *et al.* (2016a) ‘Highly aligned stromal collagen is a negative prognostic factor following pancreatic ductal adenocarcinoma resection’, *Oncotarget*, 7(46), p. In preparation. doi: 10.18632/oncotarget.12772.
Drifka, C. R. *et al.* (2016b) ‘Highly aligned stromal collagen is a negative prognostic factor following pancreatic ductal adenocarcinoma resection’, *Oncotarget*, 7(46), p. In preparation. doi: 10.18632/oncotarget.12772.
Driskell, R. R. *et al.* (2013) ‘Distinct fibroblast lineages determine dermal architecture in skin development and repair’, *Nature*, 504(7479), pp. 277–281. doi: 10.1038/nature12783.
Duclos, G. *et al.* (2014) ‘Perfect nematic order in confined monolayers of spindle-shaped cells’, *Soft Matter*. doi: 10.1039/c3sm52323c.
Egeblad, M. *et al.* (2008) ‘Visualizing stromal cell dynamics in different tumor microenvironments by spinning disk confocal microscopy’, *Disease Models and Mechanisms*, 1(2–3), pp. 155–167. doi: 10.1242/dmm.000596.
El-Naggar, A. M. *et al.* (2015) ‘Translational Activation of HIF1α by YB-1 Promotes Sarcoma Metastasis’, *Cancer Cell*. doi: 10.1016/j.ccell.2015.04.003.
Ellery, A. J. *et al.* (2016) ‘Modeling transport through an environment crowded

by a mixture of obstacles of different shapes and sizes', *Physica A: Statistical Mechanics and its Applications*, 449, pp. 74–84. doi: 10.1016/j.physa.2015.12.123.
Elsdale, T. R. (1968) 'Parallel orientation of fibroblasts in vitro', *Experimental Cell Research*, 51(2–3), pp. 439–450. doi: 10.1016/0014-4827(68)90134-1.
Escaff, D. *et al.* (2018) 'A continuous-time persistent random walk model for flocking', *Chaos*. doi: 10.1063/1.5027734.
Ewels, P. A. *et al.* (2019) 'nf-core: Community curated bioinformatics pipelines', *bioRxiv*, p. 610741. doi: 10.1101/610741.
Feig, C. *et al.* (2013) 'Targeting CXCL12 from FAP-expressing carcinoma-associated fibroblasts synergizes with anti – PD-L1 immunotherapy in pancreatic cancer', *Proc Natl Acad Sci U S A*, 110(50), pp. 20212–20217. doi: 10.1073/pnas.1320318110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1320318110.
Franco-Barraza, J. *et al.* (2016) 'Preparation of Extracellular Matrices Produced by Cultured and Primary Fibroblasts', *Current protocols in cell biology*, 71, pp. 10.9.1-10.9.34. doi: 10.1002/cpcb.2.
Frantz, C., Stewart, K. M. and Weaver, V. M. (2010) 'The extracellular matrix at a glance', *Journal of Cell Science*. doi: 10.1242/jcs.023820.
Friedl, P. *et al.* (2012) 'Classifying collective cancer cell invasion', *Nature Cell Biology*, 14(8), pp. 777–783. doi: 10.1038/ncb2548.
Gatenby, R. A. and Gawlinski, E. T. (2003) 'The glycolytic phenotype in carcinogenesis and tumor invasion: Insights through mathematical models', *Cancer Research*, 63(14), pp. 3847–3854.
Ginelli, F. *et al.* (2010) 'Large-scale collective properties of self-propelled rods', *Physical Review Letters*. doi: 10.1103/PhysRevLett.104.184502.
Goetz, J. G. *et al.* (2011) 'Biomechanical remodeling of the microenvironment by stromal caveolin-1 favors tumor invasion and metastasis', *Cell*, 146(1), pp. 148–163. doi: S0092-8674(11)00645-3 [pii]\r10.1016/j.cell.2011.05.040.
Gorelik, R. and Gautreau, A. (2014) 'Quantitative and unbiased analysis of directional persistence in cell migration.', *Nature protocols*. Nature Publishing Group, 9(8), pp. 1931–43. doi: 10.1038/nprot.2014.131.

Gotwals, P. *et al.* (2017) ‘Prospects for combining targeted and conventional cancer therapy with immunotherapy’, *Nat Rev Cancer*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 17(5), pp. 286–301. Available at: http://dx.doi.org/10.1038/nrc.2017.17.
Graner, F. and Glazier, J. A. (1992) ‘Simulation of Biological Cell Sorting Using a 2-Dimensional Extended Potts-Model’, *Physical Review Letters*, pp. 2013–2016. doi: 10.1103/PhysRevLett.69.2013.
Grégoire, G., Chaté, H. and Tu, Y. (2003) ‘Moving and staying together without a leader’, *Physica D: Nonlinear Phenomena*. doi: 10.1016/S0167-2789(03)00102-7.
Grossman, D., Aranson, I. S. and Ben Jacob, E. (2008) ‘Emergence of agent swarm migration and vortex formation through inelastic collisions’, *New Journal of Physics*. doi: 10.1088/1367-2630/10/2/023036.
Heindl, A., Nawaz, S. and Yuan, Y. (2015) ‘Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology’, *Laboratory Investigation*, 95, pp. 377–384. doi: 10.1038/labinvest.2014.155.
van Helvert, S., Storm, C. and Friedl, P. (2018) ‘Mechanoreciprocity in cell migration’, *Nature Cell Biology*. Springer US, 20(1), pp. 8–20. doi: 10.1038/s41556-017-0012-0.
Hirata, E. *et al.* (2015) ‘Intravital imaging reveals how BRAF inhibition generates drug-tolerant microenvironments with high integrin β1/FAK Signaling’, *Cancer Cell*. doi: 10.1016/j.ccell.2015.03.008.
Hirata, E. and Sahai, E. (2017) ‘Tumor microenvironment and differential responses to therapy’, *Cold Spring Harbor Perspectives in Medicine*. doi: 10.1101/cshperspect.a026781.
Hirst, L. S. and Charras, G. (2014) ‘Biological physics: Liquid crystals in living tissue’, *Nature*, 544(7649), pp. 164–165. doi: 10.1038/544164a.
Jamal-Hanjani, M. *et al.* (2017) ‘Tracking the Evolution of Non–Small-Cell Lung Cancer’, *New England Journal of Medicine*. doi: 10.1056/nejmoa1616288.
Jiao, Y. and Torquato, S. (2011) ‘Emergent behaviors from a cellular automaton model for invasive tumor growth in heterogeneous microenvironments’, *PLoS Computational Biology*, 7(12), pp. 10–12. doi: 10.1371/journal.pcbi.1002314.

Jin, C. *et al.* (2016) 'The efficacy and safety of nivolumab in the treatment of advanced melanoma: A meta-analysis of clinical trials', *OncoTargets and Therapy*, 9, pp. 1571–1578. doi: 10.2147/OTT.S96762.
Joyce, J. A. and Fearon, D. T. (2015) 'T cell exclusion, immune privilege, and the tumor microenvironment', *Science*, 348(6230), pp. 74–80. doi: 10.1126/science.aaa6204.
Kalluri, R. (2016) 'The biology and function of fibroblasts in cancer', *Nature Reviews Cancer*. Nature Publishing Group, 16(9), pp. 582–598. doi: 10.1038/nrc.2016.73.
Kalluri, R. and Zeisberg, M. (2006) 'Fibroblasts in cancer', *Nature reviews. Cancer*, 6(5), pp. 392–401. doi: 10.1038/nrc1877.
Karagiannis, G. S. *et al.* (2012) 'Cancer-Associated Fibroblasts Drive the Progression of Metastasis through both Paracrine and Mechanical Pressure on Cancer Tissue', *Molecular Cancer Research*, 10(11), pp. 1403–1418. doi: 10.1158/1541-7786.MCR-12-0307.
Karagiannis, G. S. *et al.* (2014) 'Collective migration of cancer-associated fibroblasts is enhanced by overexpression of tight junction-associated proteins claudin-11 and occludin', *Molecular Oncology*. Elsevier B.V, 8(2), pp. 178–195. doi: 10.1016/j.molonc.2013.10.008.
Kaur, A. *et al.* (2019) 'Remodeling of the collagen matrix in aging skin promotes melanoma metastasis and affects immune cell motility', *Cancer Discovery*. doi: 10.1158/2159-8290.CD-18-0193.
Keren, K. *et al.* (2008) 'Mechanism of shape determination in motile cells', *Nature*. doi: 10.1038/nature06952.
Kim, M. C. *et al.* (2015) 'Cell Invasion Dynamics into a Three Dimensional Extracellular Matrix Fibre Network', *PLoS Computational Biology*, 11(10), pp. 1–29. doi: 10.1371/journal.pcbi.1004535.
Labernadie, A. *et al.* (2017) 'A mechanically active heterotypic E-cadherin/N-cadherin adhesion enables fibroblasts to drive cancer cell invasion', *Nature Cell Biology*, 19(3), pp. 224–237. doi: 10.1038/ncb3478.
LeBleu, V. S. and Kalluri, R. (2018) 'A peek into cancer-associated fibroblasts: origins, functions and translational impact', *Disease Models & Mechanisms*. doi:

10.1242/dmm.029447.
Lecaudey, V. and Gilmour, D. (2006) ‘Organizing moving groups during morphogenesis’, *Current Opinion in Cell Biology*, 18(1), pp. 102–107. doi: 10.1016/j.ceb.2005.12.001.
Lecun, Y., Bengio, Y. and Hinton, G. (2015) ‘Deep learning’, *Nature*. doi: 10.1038/nature14539.
Leung, S. J., Rice, P. S. and Barton, J. K. (2015) ‘In vivo molecular mapping of the tumor microenvironment in an azoxymethane-treated mouse model of colon carcinogenesis’, *Lasers in Surgery and Medicine*, 47(1), pp. 40–49. doi: 10.1002/lsm.22309.
Levental, K. R. *et al.* (2009) ‘Matrix Crosslinking Forces Tumor Progression by Enhancing Integrin Signaling’, *Cell*, 139(5), pp. 891–906. doi: 10.1016/j.cell.2009.10.027.
Li, X. *et al.* (2017) ‘On the mechanism of long-range alignment order of fibroblasts’, *bioRxiv*, 119669.
Lindahl, T. (1974) ‘An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues.’, *Proceedings of the National Academy of Sciences of the United States of America*, 71(9), pp. 3649–53. doi: 10.1073/pnas.71.9.3649.
Lukeman, R., Li, Y.-X. and Edelstein-Keshet, L. (2010) ‘Inferring individual rules from collective behavior’, *Proceedings of the National Academy of Sciences*, 107(28), pp. 12576–12580. doi: 10.1073/pnas.1001763107.
Lynch, S. E. *et al.* (1987) ‘Role of platelet-derived growth factor in wound healing: synergistic effects with other growth factors.’, *Proceedings of the National Academy of Sciences of the United States of America*, 84(November), pp. 7696–7700.
Madsen, C. D. *et al.* (2015) ‘Hypoxia and loss of PHD 2 inactivate stromal fibroblasts to decrease tumour stiffness and metastasis’, *EMBO reports*, 16(10), pp. 1394–1408. doi: 10.15252/embr.201540107.
Maiuri, P. *et al.* (2015) ‘Actin flows mediate a universal coupling between cell speed and cell persistence’, *Cell*, 161(2), pp. 374–386. doi: 10.1016/j.cell.2015.01.056.

Maley, C. C. *et al.* (2017) ‘Classifying the evolutionary and ecological features of neoplasms’, *Nature Reviews Cancer*. doi: 10.1038/nrc.2017.69.
Manning, C. S. *et al.* (2013) ‘Intravital imaging reveals conversion between distinct tumor vascular morphologies and localized vascular response to Sunitinib’, *IntraVital*. doi: 10.4161/intv.24790.
Marchetti, M. C. *et al.* (2013) ‘Hydrodynamics of soft active matter’, *Reviews of Modern Physics*. doi: 10.1103/RevModPhys.85.1143.
Marée, A. F. M., Grieneisen, V. A. and Edelstein-Keshet, L. (2012) ‘How cells integrate complex stimuli: The effect of feedback from phosphoinositides and cell shape on cell polarization and motility’, *PLoS Computational Biology*, 8(3). doi: 10.1371/journal.pcbi.1002402.
Matzavinos, A. (2004) ‘Mathematical modelling of the spatio-temporal response of cytotoxic T-lymphocytes to a solid tumour’, *Mathematical Medicine and Biology*, 21(1), pp. 1–34. doi: 10.1093/imammb/21.1.1.
Mayor, R. and Carmona-Fontaine, C. (2010) ‘Keeping in touch with contact inhibition of locomotion’, *Trends in Cell Biology*. Elsevier Ltd, 20(6), pp. 319–328. doi: 10.1016/j.tcb.2010.03.005.
Mayor, R. and Etienne-Manneville, S. (2016) ‘The front and rear of collective cell migration’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 17(2), pp. 97–109. doi: 10.1038/nrm.2015.14.
Mayorca-Guiliani, A. E. *et al.* (2017) ‘ISDoT: In situ decellularization of tissues for high-resolution imaging and proteomic analysis of native extracellular matrix’, *Nature Medicine*. doi: 10.1038/nm.4352.
McDougall, S. *et al.* (2006) ‘Fibroblast migration and collagen deposition during dermal wound healing: mathematical modelling and clinical implications.’, *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 364(1843), pp. 1385–405. doi: 10.1098/rsta.2006.1773.
McGranahan, N. and Swanton, C. (2015) ‘Biological and therapeutic impact of intratumor heterogeneity in cancer evolution’, *Cancer Cell*, 27(1), pp. 15–26. doi: 10.1016/j.ccell.2014.12.001.
McGranahan, N. and Swanton, C. (2017) ‘Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future’, *Cell*, pp. 613–628. doi:

10.1016/j.cell.2017.01.018.
Meinhardt, H. (1999) ‘Orientation of chemotactic cells and growth cones: models and mechanisms.’, *Journal of cell science*, 112 (Pt 1, pp. 2867–2874.
Metzner, C. *et al.* (2015) ‘Superstatistical analysis and modelling of heterogeneous random walks’, *Nature Communications*. doi: 10.1038/ncomms8516.
Mogilner, A., Allard, J. and Wollman, R. (2012) ‘Cell polarity: quantitative modeling as a tool in cell biology’, *Science*, 336(6078), pp. 175–179. doi: 10.1126/science.1216380.
Natarajan, S. *et al.* (2019) ‘Collagen remodeling in the hypoxic tumor-mesothelial niche promotes ovarian cancer metastasis’, *Cancer Research*. doi: 10.1158/0008-5472.CAN-18-2616.
Nawaz, S. *et al.* (2015) ‘Beyond immune density: Critical role of spatial heterogeneity in estrogen receptor-negative breast cancer’, *Modern Pathology*. doi: 10.1038/modpathol.2015.37.
Newman, T. J. (2005) ‘Modeling multicellular systems using subcellular elements.’, *Mathematical biosciences and engineering : MBE*, 2(3), pp. 613–624. doi: 10.3934/mbe.2005.2.613.
Notbohm, J. *et al.* (2016) ‘Cellular Contraction and Polarization Drive Collective Cellular Motion’, *Biophysical Journal*, 110(12), pp. 2729–2738. doi: 10.1016/j.bpj.2016.05.019.
Obert, M., Pfeifer, P. and Sernetz, M. (1990) ‘Microbial growth patterns described by fractal geometry’, *Journal of Bacteriology*, 172(3), pp. 1180–1185. doi: 10.1128/jb.172.3.1180-1185.1990.
Öhlund, D. *et al.* (2017) ‘Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer’, *The Journal of Experimental Medicine*, 214(3), pp. 579 LP – 596. Available at: http://jem.rupress.org/content/214/3/579.abstract.
Osborne, J. M. *et al.* (2017) ‘Comparing individual-based approaches to modelling the self-organization of multicellular tissues’, *PLoS Computational Biology*. doi: 10.1371/journal.pcbi.1005387.
Özdemir, B. C. *et al.* (2014) ‘Depletion of carcinoma-associated fibroblasts and

fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival', *Cancer Cell*, 25(6), pp. 719–734. doi: 10.1016/j.ccr.2014.04.005.
De Palo, G., Yi, D. and Endres, R. G. (2017) 'A critical-like collective state leads to long-range cell communication in Dictyostelium discoideum aggregation', *PLoS Biology*, 15(4), pp. 1–25. doi: 10.1371/journal.pbio.1002602.
Paluch, E. K. *et al.* (2015) 'Mechanotransduction: use the force(s).', *BMC biology*, 13, p. 47. doi: 10.1186/s12915-015-0150-4.
Park, D. *et al.* (2019) 'Extracellular matrix anisotropy is determined by TFAP2C-dependent regulation of cell collisions', *Nature Materials, Accepted, in press.*
Peruani, F., Deutsch, A. and Bär, M. (2006) 'Nonequilibrium clustering of self-propelled rods', *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. doi: 10.1103/PhysRevE.74.030904.
Petrie, R. J., Doyle, A. D. and Yamada, K. M. (2009) 'Random versus directionally persistent cell migration.', *Nature reviews. Molecular cell biology*, 10(8), pp. 538–49. doi: 10.1038/nrm2729.
Pizzo, a M. *et al.* (2005) 'Extracellular matrix (ECM) microstructural composition regulates local cell-ECM biomechanics and fundamental fibroblast behavior: a multidimensional perspective.', *Journal of applied physiology (Bethesda, Md. : 1985)*, 98(5), pp. 1909–1921. doi: 10.1152/japplphysiol.01137.2004.
Provenzano, P. P. *et al.* (2006) 'Collagen reorganization at the tumor-stromal interface facilitates local invasion', *BMC Med*, 4(1), p. 38. doi: 10.1186/1741-7015-4-38.
Provenzano, P. P. *et al.* (2008) 'Collagen density promotes mammary tumor initiation and progression', *BMC Med*, 6(1), p. 11. doi: 10.1186/1741-7015-6-11.
Puliafito, A. *et al.* (2011) 'Collective and single cell behavior in epithelial contact inhibition', *Proceedings of the National Academy of Sciences of the United States of America*, 109, pp. 739–44. doi: 10.1073/pnas.1007809109.
Quaranta, V. *et al.* (2009) *Trait Variability of Cancer Cells Quantified by High-Content Automated Microscopy of Single Cells*. 1st edn, *Methods in Enzymology*. 1st edn. Elsevier Inc. doi: 10.1016/S0076-6879(09)67002-6.

Reis, J. P., Silva, E. A. and Pinho, P. (2016) ‘Spatial metrics to study urban patterns in growing and shrinking cities’, *Urban Geography*. doi: 10.1080/02723638.2015.1096118.
Rezakhaniha, R. *et al.* (2012) ‘Experimental investigation of collagen waviness and orientation in the arterial adventitia using confocal laser scanning microscopy’, *Biomechanics and Modeling in Mechanobiology*, 11(3–4), pp. 461–473. doi: 10.1007/s10237-011-0325-z.
Riching, K. M. *et al.* (2015) ‘3D collagen alignment limits protrusions to enhance breast cancer cell persistence’, *Biophysical Journal*, 107(11), pp. 2546–2558. doi: 10.1016/j.bpj.2014.10.035.
Rosser, G. *et al.* (2013) ‘The effect of sampling rate on observed statistics in a correlated random walk’, *Journal of the Royal Society Interface*. doi: 10.1098/rsif.2013.0273.
Roxanis, I. *et al.* (2018) ‘The significance of tumour microarchitectural features in breast cancer prognosis: A digital image analysis’, *Breast Cancer Research*. doi: 10.1186/s13058-018-0934-x.
Roycroft, A. and Mayor, R. (2015) ‘Forcing contact inhibition of locomotion’, *Trends in Cell Biology*, pp. 373–375. doi: 10.1016/j.tcb.2015.05.001.
Sadati, M. *et al.* (2013) ‘Collective migration and cell jamming’, *Differentiation*, 86(3), pp. 121–125. doi: 10.1016/j.diff.2013.02.005.
Sánchez-Gutiérrez, D. *et al.* (2015) ‘Fundamental physical cellular constraints drive self-organization of tissues’, *The EMBO journal*, 35(1), pp. 1–12. doi: 10.15252/embj.201592374.
Sandersius, S. a and Newman, T. J. (2008) ‘Modeling cell rheology with the Subcellular Element Model.’, *Physical biology*, 5(1), p. 015002. doi: 10.1088/1478-3975/5/1/015002.
Sandersius, S. a, Weijer, C. J. and Newman, T. J. (2011) ‘Emergent cell and tissue dynamics from subcellular modeling of active biomechanical processes.’, *Physical biology*, 8(4), p. 045007. doi: 10.1088/1478-3975/8/4/045007.
Sartakhti, J. S. *et al.* (2016) ‘Evolutionary Dynamics of Tumor-Stroma Interactions in Multiple Myeloma’, *PLoS ONE*. Edited by J. A. R. Marshall. San Francisco, CA USA: Public Library of Science, 11(12), p. e0168856. doi:

10.1371/journal.pone.0168856.
Saw, T. B. *et al.* (2017) 'Topological defects in epithelia govern cell death and extrusion', *Nature*. Macmillan Publishers Limited, part of Springer Nature. All rights reserved., 544(7649), pp. 212–216. Available at: http://dx.doi.org/10.1038/nature21718.
Schlüter, D. K., Ramis-Conde, I. and Chaplain, M. A. J. (2012) 'Computational modeling of single-cell migration: The leading role of extracellular matrix fibers', *Biophysical Journal*, 103(6), pp. 1141–1151. doi: 10.1016/j.bpj.2012.07.048.
Schumacher, L. J., Maini, P. K. and Baker, R. E. (2017) 'Semblance of Heterogeneity in Collective Cell Migration', *Cell Systems*. Elsevier. doi: 10.1016/j.cels.2017.06.006.
Scianna, M. and Preziosi, L. (2013) *Cellular Potts Models: Multiscale Extensions and Biological Applications*. CRC Press (Chapman & Hall/CRC Mathematical and Computational Biology). Available at: https://books.google.co.uk/books?id=jEDSBQAAQBAJ.
Scott, J. G. *et al.* (2016) 'Spatial Metrics of Tumour Vascular Organisation Predict Radiation Efficacy in a Computational Model', *PLoS Computational Biology*. doi: 10.1371/journal.pcbi.1004712.
Seppa, H. *et al.* (1982) 'Platelet-derived growth factor is chemotactic for fibroblasts', *Journal of Cell Biology*, 92(2), pp. 584–588. doi: 10.1083/jcb.92.2.584.
Sepúlveda, N. *et al.* (2013) 'Collective Cell Motion in an Epithelial Sheet Can Be Quantitatively Described by a Stochastic Interacting Particle Model', *PLoS Computational Biology*, 9(3). doi: 10.1371/journal.pcbi.1002944.
Shiga, K. *et al.* (2015) 'Cancer-associated fibroblasts: Their characteristics and their roles in tumor growth', *Cancers*, pp. 2443–2458. doi: 10.3390/cancers7040902.
Shimoda, M., Mellody, K. T. and Orimo, A. (2010) 'Carcinoma-associated fibroblasts are a rate-limiting determinant for tumour progression', *Seminars in Cell and Developmental Biology*. Elsevier Ltd, 21(1), pp. 19–25. doi: 10.1016/j.semcdb.2009.10.002.
Stacklies, W. *et al.* (2007) 'pcaMethods—a bioconductor package providing

PCA methods for incomplete data', *Bioinformatics*, 23(9), pp. 1164–1167. Available at: http://dx.doi.org/10.1093/bioinformatics/btm069.
Steger, G. (1998) 'An unbiased detector of curvilinear structures', *IEEE Transactions on Pattern Analysis and Machine Intelligence*. doi: 10.1109/34.659930.
Steinwachs, J. *et al.* (2015) 'Three-dimensional force microscopy of cells in biopolymer networks', *Nature Methods*, 13(2), pp. 171–176. doi: 10.1038/nmeth.3685.
Stramer, B. and Mayor, R. (2016) 'Mechanisms and in vivo functions of contact inhibition of locomotion', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, (Cil). doi: 10.1038/nrm.2016.118.
Swanson, K. R. *et al.* (2003) 'Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion', *Journal of the Neurological Sciences*, 216(1), pp. 1–10. doi: 10.1016/j.jns.2003.06.001.
Szabó, A. and Mayor, R. (2016) 'Modelling collective cell migration of neural crest', *Current Opinion in Cell Biology*, 42, pp. 22–28. doi: 10.1016/j.ceb.2016.03.023.
Szabó, B. *et al.* (2006) 'Phase transition in the collective migration of tissue cells: Experiment and model', *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 74(6). doi: 10.1103/PhysRevE.74.061908.
Tozluoğlu, M. *et al.* (2013a) 'Matrix geometry determines optimal cancer cell migration strategy and modulates response to interventions.', *Nature cell biology*. Nature Publishing Group, 15(7), pp. 1–14. doi: 10.1038/ncb2775.
Tozluoğlu, M. *et al.* (2013b) 'Matrix geometry determines optimal cancer cell migration strategy and modulates response to interventions', *Nature Cell Biology*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 15, p. 751. Available at: https://doi.org/10.1038/ncb2775.
Trepat, X. *et al.* (2009) 'Physical forces during collective cell migration', *Nature Physics*. Nature Publishing Group, 5(6), pp. 426–430. doi: 10.1038/nphys1269.
Trepat, X. and Fredberg, J. J. (2011) 'Plithotaxis and emergent dynamics in collective cellular migration', *Trends in Cell Biology*, pp. 638–646. doi: 10.1016/j.tcb.2011.06.006.

Tweedy, L. *et al.* (2013) 'Distinct cell shapes determine accurate chemotaxis', *Scientific Reports*, 3, pp. 1–7. doi: 10.1038/srep02606.
Valastyan, S. and Weinberg, R. A. (2011) 'Tumor metastasis: Molecular insights and evolving paradigms', *Cell*. Elsevier Inc., 147(2), pp. 275–292. doi: 10.1016/j.cell.2011.09.024.
Vedula, S. R. K. *et al.* (2012) 'Emerging modes of collective cell migration induced by geometrical constraints.', *Proceedings of the National Academy of Sciences of the United States of America*, 109(32), pp. 12974–9. doi: 10.1073/pnas.1119313109.
Vicsek, T. *et al.* (1995) 'Novel type of phase transition in a system of self-driven particles', *Physical Review Letters*, 75(6), pp. 1226–1229. doi: 10.1103/PhysRevLett.75.1226.
Vicsek, T. and Zafiris, A. (2010) 'Collective motion', p. 52. doi: 10.1016/j.physrep.2012.03.004.
Wang, K. *et al.* (2016) 'Hidden in the mist no more : physical force in cell biology', *Nature Publishing Group*. Nature Publishing Group, 13(2), pp. 124–125. doi: 10.1038/nmeth.3744.
Wershof, E. *et al.* (2019) 'Matrix feedback enables diverse higher-order patterning of the extracellular matrix', *PLoS Computational Biology, Accepted in press.*
Wilkins, A. *et al.* (2018) 'Methodology for tissue sample collection within a translational sub-study of the CHHiP trial (CRUK/06/016), a large randomised phase III trial in localised prostate cancer', *Clinical and Translational Radiation Oncology*. doi: 10.1016/j.ctro.2018.02.002.
Wilkins, A. C. *et al.* (2018) 'Ki67 Is an Independent Predictor of Recurrence in the Largest Randomized Trial of 3 Radiation Fractionation Schedules in Localized Prostate Cancer', *International Journal of Radiation Oncology Biology Physics*. doi: 10.1016/j.ijrobp.2018.01.072.
Woods, M. L. *et al.* (2014) 'Directional collective cell migration emerges as a property of cell interactions', *PLoS ONE*, 9(9). doi: 10.1371/journal.pone.0104969.
Wu, A. *et al.* (2014) 'Game theory in the death galaxy: interaction of cancer and

stromal cells in tumour microenvironment.', *Interface focus*, 4(4), p. 20140028. doi: 10.1098/rsfs.2014.0028.
Wu, P.-H. *et al.* (2014) 'Three-dimensional cell migration does not follow a random walk.', *Proceedings of the National Academy of Sciences of the United States of America*, 111(11), pp. 3949–54. doi: 10.1073/pnas.1318967111.
Yin, Z. *et al.* (2013) 'A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes.', *Nature cell biology*. Nature Publishing Group, 15(7), pp. 860–71. doi: 10.1038/ncb2764.
Yolland, L. *et al.* (2019) 'Persistent and polarized global actin flow is essential for directionality during cell migration', *Nature Cell Biology*. doi: 10.1038/s41556-019-0411-5.
Yuan, Y. (2016) 'Spatial heterogeneity in the tumor microenvironment', *Cold Spring Harbor Perspectives in Medicine*, 6(8). doi: 10.1101/cshperspect.a026583.
Zaritsky, A. *et al.* (2015) 'Seeds of Locally Aligned Motion and Stress Coordinate a Collective Cell Migration', *Biophysical Journal*, 109(12), pp. 2492–2600. doi: 10.1016/j.bpj.2015.11.001.
Zimmermann, J. *et al.* (2016) 'Contact inhibition of locomotion determines cell-cell and cell-substrate forces in tissues', *Proceedings of the National Academy of Sciences of the United States of America*, 113(10), pp. 2660–2665. doi: 10.1073/pnas.1522330113.

Chapter 6. Appendix

6.1 Statistical test for collision guidance success

The statistical test used is a z-test using a pooled estimate of variance. We want to test if two binomial distributions are significantly different from each other. This can be constructed as the following hypothesis test:

$$H_0: p_1 = p_2, \qquad H_1: p_1 > p_2$$

where p_i is the probability of success in the binomial distribution $X \sim Bin(n_i, p_i)$. We calculate the test statistic as

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}$$

where $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$, and compare this statistic against the critical region value of $z_c = 1.645$ defining a one-tailed test with significance if p<0.05. We reject H_0 if $z > z_c$.

For all other hypothesis testing, we used one or two-tailed t-tests. One-tailed t-tests were used when the hypotheses were:

$$H_0: p_1 = p_2, \qquad H_1: p_1 > p_2$$

and two-tailed t-tests were used when

$$H_0: p_1 = p_2, \qquad H_1: p_1 \neq p_2$$

6.2 Experimental methods

Experimental methods are given for interest but were written by Danielle Park and not by the author.

6.2.1 Cell lines and reagents

Aligning VCAF8 fibroblasts were isolated from a human vulval carcinoma and immortalised with lentiviral HTERT. Cells were selected using 400 µg mL-1 hygromycin. Non-Aligning CAF1 were isolated from transgenic FVB/n mice expressing the Polyoma Middle T antigen oncogene under the Mouse Mammary Tumour Virus promoter (MMTV-PyMT) as described in (Calvo *et al.*, 2013). Cells were immortalised with HPV-E6 retrovirus and selected using 2.5 µg mL-1 puromycin. Cells were maintained in DMEM (Invitrogen), 10% FCS (PAA Labs), 1% ITS (insulin–transferrin–selenium; #41400-045; Invitrogen) supplement. For persistence analysis cells were infected with the retroviral nuclear tag AcGFP-NLS (pLNCX2) and selected using 500 µg mL-1 geneticin. To inhibit persistence cells were treated with 0.5µM PDGF tyrosine kinase inhibitor IV (Merck, 521233) and tracked for 16h. For co-culture experiments VCAF8 and CAF1 were infected with the lentiviral membrane tag GFP-CAAX and mcherry-CAAX respectively (pCSII-IRES2).

Human fibroblasts were isolated from patient tissues of vulval (VCAF8) and immortalised with lentiviral HTERT as described in Gaggioli, C., et al. All patient samples were collected under ethical approval 10/H0304/14 and 15/EE/0151. Cells were selected using 400 µg mL-1 hygromycin and maintained in DMEM (Invitrogen), 10% FCS (PAA Labs), 1% ITS (insulin–transferrin–selenium; #41400-045; Invitrogen) supplement.

Mouse fibroblasts (CAF1) were isolated from transgenic FVB/n mice expressing the Polyoma Middle T antigen oncogene under the Mouse Mammary Tumour Virus promoter (MMTV-PyMT) as described in Calvo, F., et al. Cells were immortalised

with HPV-E6 retrovirus, selected using 2.5 µg mL-1 puromycin and maintained in DMEM, 10% FCS and 1% ITS as above.

For cell tracking experiments fibroblasts were infected with the retroviral nuclear tag AcGFP-NLS (pLNCX2) and selected using 500 µg mL-1 geneticin. The functional upstream domain (FUD) of adhesin F1 of *Streptococcus pyogene* was used to disrupt fibronectin assembly and was kindly provided by C. Albiges-Rizo (IAB, Grenoble, France).

6.2.2 Immunofluorescence

1 x 10^5 cells were seeded on a gelatin coated glass bottom 24 well MatTek dish in DMEM 10% FBS, 1% ITS, 50 ug/ml ascorbic acid ((+)-Sodium L-ascorbate, A4034, Sigma) and allowed to grow to confluence over 7 days as described in (Franco-Barraza *et al.*, 2016). Media was replaced every second day with fresh ascorbic acid. Cells were fixed in 4% paraformaldehyde for 20 min, washed once with PBS and blocked for 60 min at room temperature (RT) in blocking solution: 4% BSA PBS 0.05% Tween20. Cells were then incubated with a rabbit anti-fibronectin antibody (Sigma, F3648) in blocking solution overnight at 4 °C. After 3 washes of 15 min in PBS, the secondary donkey anti-rabbit antibody Alexa Fluor 555 (Invitrogen, A31572) was added in blocking solution. After 3 washes of 15 min in PBS, cells were permeabilised by incubation with PBS 0.2% Triton 100 (Sigma) at RT for 20 min, followed by incubation with Phalloidin 633 (Invitrogen, 68825) and DAPI to visualise F-actin and nuclei respectively. Cells were imaged and analysed using an inverted Zeiss LSM880 confocal microscope.

6.2.3 Phase contrast time-lapse imaging

1 x 10^4 CAFs were seeded onto a 24 well glass bottom MatTek dish and imaged 8 hrs later. Bright-field and epifluorescence time-lapse imaging was performed at 37 °C and 5% CO_2 with an inverted microscope (Nikon ECLIPSE TE2000-E). Bright-field and epifluorescence images were taken every 5 min through a ×10 PlanFluor,

NA 0.3 Ph1, Nikon objective. The imaging system includes a Xenon PE300BF lamp, an Andor iXonEM+ DU-888 back-illuminated EMCCD scientific camera, and System Control Software MetaMorph Version 7.7.3.0 (Molecular Devices).

For analysis of cell body alignment over time, images were acquired over 7 days as cells moved to confluence. For persistence analysis, images were acquired over 24 hrs to ensure cells were maintained at sub-confluence. Cell tracking was performed by tracking the GFP-positive nuclei using MetaMorph software.

6.2.4 Fibroblast derived matrix assay

The fibroblast derived matrix assay was performed as described in Franco-Barraza, J., et al. Briefly, 24 well glass bottom MatTek dishes (P35-1.5-14-C, MatTek Co., Ashland, MA, USA) were pre-prepared with 0.2% gelatin solution for 1 hr at 37 °C, followed by 1% glutaraldehyde for 30 min at room temperature. The plate was washed twice with PBS then incubated with 1M ethanolamine for 30 min at room temperature. The plate was washed twice with PBS before seeding 7 x 10^4 cells in media supplemented with 100µg/ml ascorbic acid ((+)-Sodium L-ascorbate, A4034, Sigma). The cells were maintained for 6 days and the media changed every two days. Cells were removed using the extraction buffer described and washed several times with PBS before undertaking immunofluorescence for ECM components. The ECM was stained with the anti-fibronectin antibody (1:1000 dilution, Sigma, F3648) or anti-fibronectin-FITC (1:50 dilution: Abcam, ab72686). Where indicated 7 x 10^3 fibroblasts (sub-confluent) or 7 x 10^4 cells (confluent) were plated on top of pre-existing matrices for time-lapse imaging.

6.2.5 Time-lapse microscopy for persistence analysis

Nuclear labelled fibroblasts were seeded at approx. 7 x 10^3 cells per well in 24 well glass bottom MatTek dish and imaged approximately 8 hrs later. Bright-field and epifluorescence time-lapse imaging was performed at 37 °C and 5% CO_2 with an inverted microscope (Nikon Ti2 inverted microscope fitted with a Okolab environmental chamber and CO2 mixer). Bright-field and epifluorescence images

were taken every 10 min through a ×10 PlanFluor, NA 0.3 Ph1, Nikon objective. The imaging system includes a SpectraX LED light engine (Lumencor) fitted with standard filters and Photometrics Prime scientific CMOS camera. The microscope was managed using Micro-Manager v2.0 software.

Where indicated cells were pre-treated with 500nM FUD and the media changed every 24 hrs. Cells tracked using the ImageJ Trackmate plug in. Persistence was calculated as the ratio of shortest linear distance between two points of migration (displacement) to the total distance traversed by the cell (distance) over 16 hr intervals.

6.2.6 Immunohistochemistry of breast cancer microarray

Human invasive ductal and lobular breast carcinoma microarray was stained for picrosirus red.

6.2.7 *In vivo* imaging of collagen

The organs of platelet derived growth factor receptor, alpha polypeptide; targeted mutation 11 (MGI:2663656) with nuclear labelled EGFP were imaged using second harmonic confocal microscopy. Supplementary Fig 8a shows a subsection of stomach dermis together with an *in silico* representation of a similar system. Imaging of this kind shows how matrix can change over time, for instance in ageing (Supplementary Fig 8b and 8c). Supplementary Fig 8d shows the complementary Gomori Trichome imaging of Fig 4d.

6.2.8 Fibroblast derived matrix assay

The fibroblast derived matrix assay was performed as described in Franco-Barraza, J., et al. Briefly, 24 well glass bottom MatTek dishes (P35-1.5-14-C, MatTek Co., Ashland, MA, USA) were pre-prepared with 0.2% gelatin solution for 1 hr at 37 °C, followed by 1% glutaraldehyde for 30 min at room temperature. The plate was washed twice with PBS then incubated with 1M ethanolamine for 30 min at room temperature. The plate was

washed twice with PBS before seeding 7 x 10^4 cells in media supplemented with 100µg/ml ascorbic acid ((+)-Sodium L-ascorbate, A4034, Sigma). The cells were maintained for 6 days and the media changed every two days. Cells were removed using the extraction buffer described and washed several times with PBS before undertaking immunofluorescence for ECM components. The ECM was stained with the anti-fibronectin antibody (1:1000 dilution, Sigma, F3648) or anti-fibronectin-FITC (1:50 dilution: Abcam, ab72686). Where indicated 7 x 10^3 fibroblasts (sub-confluent) or 7 x 10^4 cells (confluent) were plated on top of pre-existing matrices for time-lapse imaging